Universidade de Aveiro
2016/2017

Departamento de Química

Bárbara Beatriz Pinto Macedo

Aldehyde Dehydrogenases as potential biomarkers in Myeloid Neoplasias

Aldeído desidrogenases como potenciais biomarcadores em Neoplasias Mieloides

Universidade de Aveiro
2016/2017

Departamento de Química

Bárbara Beatriz Pinto Macedo

Aldehyde Dehydrogenases as potential biomarkers in Myeloid Neoplasias

Aldeído desidrogenases como potenciais biomarcadores em Neoplasias Mieloides

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Bioquímica, realizada sob a orientação científica do Professor Doutor Francisco Manuel Lemos Amado, Professor Auxiliar com agregação ao Departamento de Química da Universidade de Aveiro, da Professora Doutora Ana Bela Ribeiro Sarmento, Professora Associada com agregação da Faculdade de Medicina da Universidade de Coimbra, e da Doutora Ana Cristina Gonçalves, Doutora Ana Cristina Gonçalves, Bolseira de Gestão em Ciência e Tecnologia da Faculdade de Medicina da Universidade de Coimbra.

o júri

presidente

Prof. Doutora Rita Maria Pinho Ferreira
Professora auxiliar da Universidade de Aveiro

Doutora Ana Cristina Gonçalves
Bolseira de Gestão em Ciência e Tecnologia da Faculdade de Medicina da Universidade de Coimbra

Prof. Doutora Maria Joana Lima Barbosa Melo
Professora auxiliar da Faculdade de Medicina da Universidade de Coimbra

agradecimentos

Embora, pela sua finalidade académica, uma dissertação seja um trabalho individual, há contributos que não podem deixar de ser referidos. Por essa razão expresso os meus sinceros agradecimentos:

Às minhas orientadoras na Faculdade de Medicina da Universidade de Coimbra, Professora Doutora Ana Bela Sarmento Ribeiro e Doutora Ana Cristina Gonçalves, pela orientação e assistência na definição deste projeto, pelo apoio e disponibilidade ao longo do seu desenvolvimento, pelas críticas e conselhos e, em especial, pela partilha de conhecimentos e experiência científica, e pelo tempo que generosamente me dedicaram.

Ao meu orientador na Universidade de Aveiro, Professor Doutor Francisco Manuel Lemos Amado, pela orientação ao longo do desenvolvimento desta tese, pela disponibilidade e apoio que sempre demonstrou, e pelos conselhos que tão generosamente partilhou.

A todos os que durante o período de realização deste trabalho foram passando pelo Laboratório de Oncologia e Hematologia (LOH) da Faculdade de Medicina da Universidade de Coimbra pelo bom ambiente que sempre proporcionaram. Agradeço especialmente à Mestre Joana Jorge e a Mestre Raquel Alves pelos conhecimentos partilhados e constante apoio que sempre demonstraram e que foi essencial para a realização desta tese.

Às melhores companheiras de casa, laboratório e da vida em geral, Beatriz e Catarina, por estarem sempre presentes e tornarem todos os momentos mais divertidos, ninguém me faz rir como vocês. Obrigada por todas as aventuras e por todos os bons momentos que passamos nos últimos 5 anos. Acima de tudo, obrigada pelo apoio nas alturas em que mais precisei de incentivo.

À Cláudia, elemento que completa o Quarteto, que mesmo longe se manteve sempre presente e disponível. Obrigada por todas as aventuras e por todos os bons momentos que partilhamos ao longo dos últimos 5 anos. Obrigada pela amizade e apoio nos momentos de maior pânico e desânimo.

À Sara e a Inês por todas as gargalhadas. Ajudaram a tornar os últimos anos memoráveis.

A todos aqueles com quem partilhei bons momento e que fizeram da minha passagem por Aveiro e Coimbra uma experiência que jamais esquecerei.

À Anne pela amizade, apoio e, principalmente, partilha de músicas que tornaram a escrita desta tese bem mais fácil. Obrigada por todas as tardes passadas na biblioteca e por todos os cafés de fim de semana. Sem ti não tinha conseguido.

À Ana pela amizade incondicional ao longo dos últimos 17 anos. Obrigada por todo o apoio e incentivo durante este ano e, principalmente, durante a vida. Obrigada por todas as horas de conversa e momentos de descontração. Acima de tudo, obrigada por seres a melhor amiga que alguém pode ter.

Por último, à minha família por todo o apoio ao longo deste ano.

Aos meus pais pelo carinho e apoio incondicional, por me encorajarem a ser sempre melhor e, acima de tudo, por acreditarem em mim.

À Carolina, a melhor e mais chata irmã do mundo, obrigada por todo o apoio e por fazeres de mim uma pessoa melhor, afinal alguém tem que dar o exemplo. A minha vida não faria sentido sem ti.

Um grande Obrigado a todos…

Palavras-Chave Aldeído desidrogenase; Aldeídos; Células Estaminais Hematopoiéticas; Síndromes Mielodisplásicas; Leucemia Mieloide Aguda; Biomarcador de Diagnóstico;

Resumo

A superfamília das desidrogenases dos aldeídos (ALDH) é constituída por 19 enzimas cuja principal função é a proteção do organismo contra aldeídos tóxicos. As ALDHs têm sido associadas ao desenvolvimento de múltiplas doenças. As síndromes mielodisplásicas (SMD) são caracterizadas por hematopoiese ineficaz associada a citopenias no sangue periférico e elevada predisposição para transformação leucémica. A leucemia mieloide aguda (LMA) é caracterizada por crescimento anómalo de células mieloides imaturas no sangue e na medula óssea. A fisiopatologia das SMDs e LMAs é um processo complexo de múltiplas etapas que envolve alterações genéticas e epigenéticas numa ampla variedade de genes associados à diferenciação, proliferação, auto-renovação e apoptose celulares. Uma vez que as ALDHs estão envolvidas em alguns destes processos biológicos, a desregulação destas enzimas pode influenciar o desenvolvimento de SMD e LMA.

O objetivo deste estudo é avaliar a expressão génica das ALDHs em doentes com SMD e LMA de modo a verificar seu potencial como biomarcadores de diagnóstico e/ou prognóstico destas doenças.

Neste contexto, analisou-se a expressão da *ALDH1A1*, *ALDH1A2*, *ALDH1A3*, *ALDH1B1*, *ALDH1L1*, *ALDH1L2*, *ALDH2*, *ALDH3A1*, *ALDH3A2*, *ALDH3B1*, *ALDH3B2*, *ALDH4A1*, *ALDH5A1*, *ALDH7A1*, *ALDH16A1* e *ALDH18A1* por PCR de transcriptase reversa. Os genes diferencialmente expressos (*ALDH3A2*, *ALDH3B1*, *ALDH4A1* e *ALDH18A1*) foram quantificados, por PCR em tempo real, em 54 doentes, 34 com SMD e 20 com LMA, e em 34 controlos saudáveis. A análise estatística foi realizada com recurso aos testes de Kolmogorov-Smirnov, Mann-Whitney, Kruskal-Wallis e análise ROC. As diferenças foram consideradas significativas quando $p<0.05$.

Os resultados mostraram que para as *ALDH1A3*, *ALDH1L1*, *ALDH1L2*, *ALDH3A1*, *ALDH3B2* e *ALDH5A1* não existem diferenças de expressão entre doentes com SMD, doentes com LMA e controlos. Os doentes com SMD e LMA apresentam níveis de expressão de *ALDH3A2* (SMD: mediana 1.9251; distância interquartil 1.28; *p*=0.000618; LMA: mediana 1.5096; distância interquartil 0.99; *p*=0.008) e *ALDH4A1* (SMD: mediana 0.1841; distância interquartil 0.47; *p*=0.01134; LMA: mediana 0.1635; distância interquartil 0.78; *p*=0.124) superiores aos observados nos controlos (*ALDH3A2*: mediana 0.4624; distância interquartil 1.53; *ALDH4A1*: mediana 0.0388; distância interquartil 0.12). Por outro lado, os doentes com SMD apresentam níveis de expressão mais elevados de *ALDH3B1* (mediana 1.6445; distância interquartil 1.39) do que os doentes com LMA (mediana 0.4541; distância interquartil 0.47; *p*=0.000314) e controlos (mediana 0.3521; distância interquartil 0.51; *p*=5.9942E-07). Os diferentes subtipos de SMD apresentam expressão diferencial de *ALDH3A2* e *ALDH3B1*. Além disso, os doentes com SMD e com LMA com alterações mielodisplásicas não expressam *ALDH18A1*. Seguidamente avaliou-se a expressão das ALDHs nos doentes de SMD estratificados de acordo com o WPSS (*WHO classification-based Prognostic Scoring System*) e verificou-se que os doentes de muito baixo risco apresentaram maior expressão de *ALDH3B1* (mediana 17.6934; distância interquartil 16.32) comparativamente aos de risco intermédio (mediana 1.2352; distância interquartil 2.55; *p*=0.010) . Por fim, a expressão das isoformas de ALDH não parece influenciar a sobrevivência global de doentes com SMD e LMA ou a evolução de SMD para LMA.

Em conclusão, este trabalho sugere que as isoformas de ALDH apresentam expressão diferencial em doentes com SMD e LMA relativamente a indivíduos saudáveis e entre si. A expressão de *ALDH3B1* poderá ser um potencial biomarcador de diagnóstico de SMD. Além disso, uma vez que nenhum doente com SMD e LMA com alterações mielodisplásicas apresenta expressão da *ALDH18A1*, a expressão desta isoforma poderá ser um bom biomarcador de diagnóstico de mielodisplasia. Por fim, a expressão de *ALDH3A2* demostrou ser um possível biomarcador de diagnóstico de LMA. No entanto, estudos adicionais com um maior número de amostras são necessários para provar o potencial dessas enzimas como biomarcadores de diagnóstico.

Keywords Aldehyde dehydrogenase; Aldehydes; Hematopoietic Stem Cells; Myelodysplastic Syndromes; Acute Myeloid Leukemia; Diagnostic Biomarkers

Abstract

Aldehyde Dehydrogenase (ALDH) superfamily is a group of 19 enzymes critical to the protection against toxic aldehydes, and have been associated with multiple diseases. Myelodysplastic syndromes (MDS) are characterized by ineffective hematopoiesis associated with peripheral blood cytopenias, and a predisposition toward leukemic transformation. Acute myeloid leukemia (AML) are characterized by disordered growth of immature myeloid blood cells in the blood and bone marrow. The pathophysiology of both, MDS and AML, is a complex multistep process that involves genetic and epigenetic abnormalities in a wide variety of genes associated with differentiation, cellular proliferation, self-renewal, and apoptosis. Since ALDHs are involved in some of these biological processes, the deregulation of these enzymes may influence MDS and AML development.

The aim of the study is to evaluate the gene expression of ALDHs in patients with MDS and AML in order to verify their potential as a biomarker for the diagnosis and/or prognosis of these diseases.

To this end, we did a preliminary analysis of the expression levels of *ALDH1A1*, *ALDH1A2*, *ALDH1A3*, *ALDH1B1*, *ALDH1L1*, *ALDH1L2*, *ALDH2*, *ALDH3A1*, *ALDH3A2*, *ALDH3B1*, *ALDH3B2*, *ALDH4A1*, *ALDH5A1*, *ALDH7A1*, *ALDH16A1*, and *ALDH18A1*. Then, we analyzed gene expression of *ALDH3A2*, *ALDH3B1*, *ALDH4A1*, and *ALDH18A1* in 54 patients, 34 MDS and 20 AML, and 34 healthy controls. ALDH expression levels were analyzed using Reverse Transcriptase-PCR and differentially expressed genes were quantified by qPCR. The statistical analysis was carried out by the Kolmogorov-Smirnov Test, Mann-Whitney test, Kruskal-Wallis test, and ROC analysis. A value of $p < 0.05$ was considered significant.

The results indicate that *ALDH1A3*, *ALDH1L1*, *ALDH1L2*, *ALDH3A1*, *ALDH3B2*, and *ALDH5A1* did not show differences in their expression between MDS, AML, and controls. *ALDH3A2*, *ALDH3B1*, *ALDH4A1*, and *ALDH18A1* had differential expression among study groups and were quantified by real time PCR. MDS and AML patients showed higher expression of *ALDH3A2* (MDS: median 1.9251; interquartile range 1.28; p=0.000618; AML: median 1.5096; interquartile range 0.99; p=0.008) and *ALDH4A*1 (MDS: median 0.1841; interquartile range 0.47; p=0.01134; AML: median 0.1635; interquartile range 0.78; p=0.124) in comparison with controls (*ALDH3A2*: median 0.4624; interquartile range 1.53; *ALDH4A1*: median 0.0388; interquartile range 0.12). On the other hand, the expression of *ALDH3B1* was higher in MDS patients (median 1.6445; interquartile range 1.39) than in AML patients (median 0.4541; interquartile range 0.47; p=0.000314) and controls (median 0.3521; interquartile range 0.51; p=5.9942E-07). *ALDH3A2* and *ALDH3B1* also showed statistically significant differences between the different subtypes of MDS. Additionally, patients with MDS and AML with myelodysplasia related changes (AML-MRC) did not expressed *ALDH18A1*. When we compared MDS patients according to WHO classification-based prognostic scoring system (WPSS) risk groups it was found that patients with very low risk had higher expression of *ALDH3B1* (median 17.6934; interquartile range 16.32) in comparison with patients with intermediate risk (median 1.2352; interquartile range 2.55; p=0.010). Furthermore, the expression of ALDH isoforms does not appear to influence MDS and AML patient's overall survival or MDS evolution to AML.

In summary, ALDH isoforms have differential expression patterns in MDS and AML patients when compared with controls and each other. The *ALDH3B1* is a potential diagnostic biomarker of MDS. Since none MDS and AML-MRC patients expressed *ALDH18A1*, the expression of this isoform may be a good diagnostic biomarker. Finally, *ALDH3A2* could be a diagnostic biomarker of AML. However, further studies with a higher number of participants are needed to prove the potential of these enzymes as diagnostic biomarkers.

Index

Index of Figures

Index of Tables

List of Abbreviations

AASA	α-Aminoadipidic-Semialdehyde
ALDH	Aldehyde Dehydrogenase
AML	Acute Myeloid Leukemia
AML-MD	Acute Myeloid Leukemia with Minimal Differentiation
AML-MRC	Acute Myeloid Leukemia with Myelodysplasia-Related Changes
AMML	Acute Myelomonocytic Leukemia
APL	Acute Promyelocytic Leukemia with PML-RARA
BMMCs	Bone Marrow Mononuclear Cells
CSCs	Cancer Stem Cells
ER	Estrogen Receptor
GSA	γ-Glutamate Semialdehyde
GSH	Reduced Glutathione
GSSG	Oxidized Glutathione
HSCs	Hematopoietic Stem Cells
MDS	Myelodysplastic Syndrome
MDS-MD	Myelodysplastic Syndrome with Multilineage Dysplasia
MDS-EB	Myelodysplastic Syndrome with Excess Blasts
MDS-RS	Myelodysplastic Syndrome with Ring Sideroblasts
MDS-SLD	Myelodysplastic Syndrome with Single Lineage Dysplasia
MPN	Myeloproliferative Neoplasm
NK	Natural Killer
P5C	Pyrroline-5-Carboxylate
P5CDH	Pyrroline-5-Carboxylate Dehydrogenase
P5CS	Δ1-Pyrroline-5-Carboxylate Synthase
PBMCs	Peripheral Blood Mononuclear Cells
RA	Retinoic Acid
RAR	Retinoic Acid Receptor
RAREs	Retinoic Acid Response Elements
RXR	Retinoid X Receptor
TICs	Tumor-Initiating Cells
t-MDS	Therapy-related Myelodysplastic Syndrome

Chapter I: Introduction

1.1 The hematopoietic system

The stem cell concept was proposed by Till and McCulloch following their pioneering studies of the blood system regeneration *in vivo* (1). In their studies, they observed the formation of cellular colonies in the spleens of recipient mice. After the transplantation of a limiting number of syngenic bone marrow cells. The analysis of these colonies revealed a small sub-population of donor bone marrow cells that possessed two extraordinary properties: ability to generate multiple types of myeloerythroid cells and to self-replicate. These findings introduced the two defining criteria of stem cells: multipotency and self-renewal (1).

Hematopoietic Stem Cells (HSCs) are the only cells in the hematopoietic system that possess the potential for both self-renewal and multipotency (Figure 1). In this context, self-renewal is the capacity to give rise to identical daughter HSCs without differentiation, while multipotency is the ability to differentiate into all functional blood cells (1,2). The HSCs exist in the bone marrow in a relatively quiescent state, thus ensuring the lifelong regeneration of the mammalian blood cells (1,2). These mature blood cells are produced at a rate of more than 1 million cells per second in the adult human (1,3) and can be of either the myeloid or lymphoid lineages. The cells of the myeloid lineage include erythrocytes, platelets, and the white blood cells of the innate immune response such as neutrophils, eosinophils, dendritic cells, and macrophages, while the lymphoid lineage produces B and T lymphocytes, and Natural Killer (NK) cells all of which are involved in the adaptive immune response (2,4)**.**

The highly dynamic and well-regulated process of self-renewal and differentiation of HSCs by which all the blood cell lineages are produced is termed hematopoiesis. The equilibrium of this process is achieved by a combination of extrinsic and intrinsic factors such as niche-associated factors, signal transduction pathways, transcription factors, and chromatin modifiers. The disruption and/or misregulation of the hematopoietic process have the potential to lead to the development of hematological disorders (2)**.**

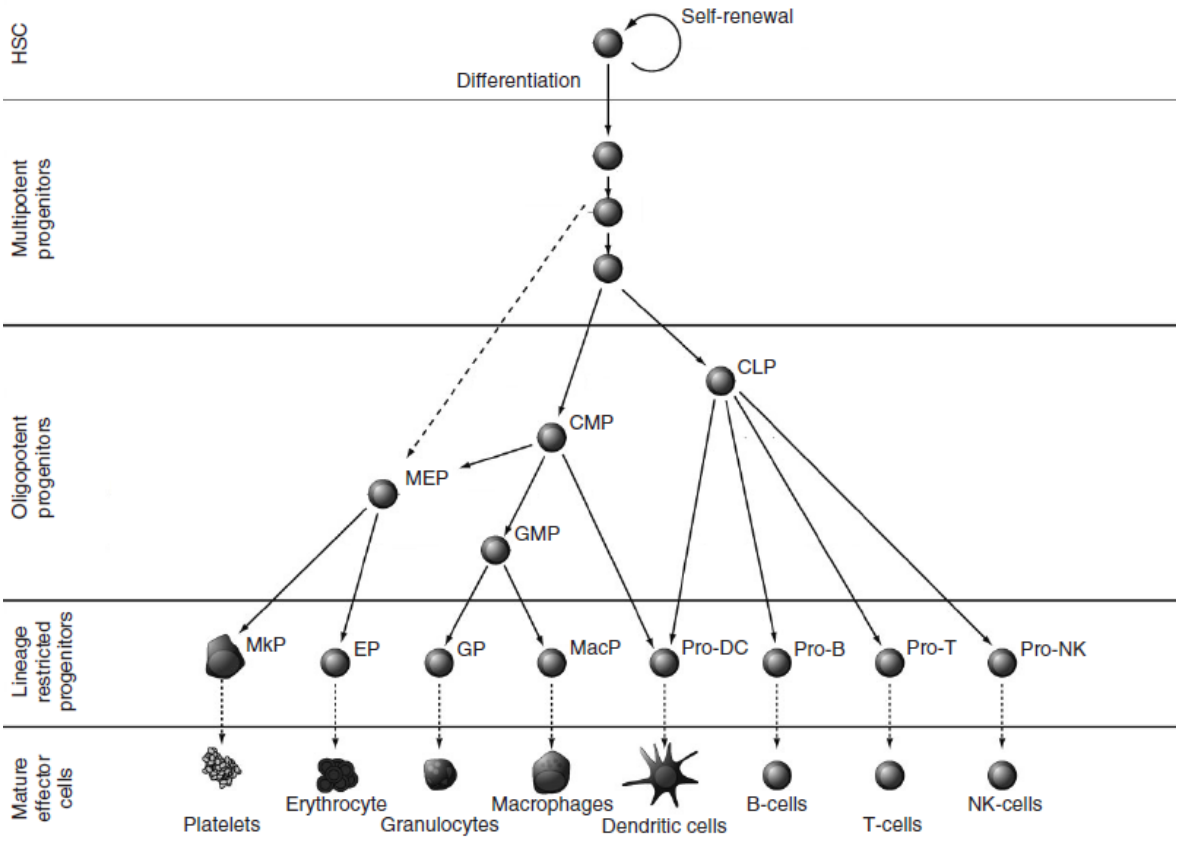

Figure 1 – Model of the hematopoietic hierarchy. The HSC is at the top of the hierarchy and is defined as the cell that has self-renewal capacity and the potential to give rise to all hematopoietic cell types. Throughout differentiation, an HSC first loses self-renewal capacity, and then loses lineage potential as it commits to become a mature functional cell of a certain lineage. Abbreviations: CLP, common lymphoid progenitor; CMP, common myeloid progenitor; DC, dendritic cell; EP, erythrocyte cell; GMP, granulocyte/macrophage progenitor; GP, granulocyte progenitor; HSC, hematopoietic stem cell; MacP, macrophage progenitor; MEP, megakaryocyte/erythrocyte progenitor; MkP, megakaryocyte progenitor; NK, natural killer. Adapted from (1).

1.2 Hematological malignancies

Hematological disorders can arise during any stage of blood cell development affecting the production and function of blood cells (2). Abnormalities in the normal hematopoietic differentiation and/or proliferation, can result in several types of blood cancers, such as leukemias, lymphomas, myelomas, and myelodysplastic syndromes.

1.2.1 Myelodysplastic Syndromes

Myelodysplastic syndromes (MDS) are a heterogeneous spectrum of clonal hematopoietic stem cell disorders characterized by ineffective hematopoiesis associated with

peripheral blood cytopenias, myelodysplasia of one, two, or all three – erythrocytic, granulocytic, and megakaryocytic – myeloid cell lineages, and genetic instability (5–7)**.** The pathophysiology of MDS is a multistep process that involves cytogenetic changes, gene mutations, or both, with a widespread gene hypermethylation (8). The deleterious nature of this disorder is underscored by its predisposition toward leukemic transformation into Acute myeloid leukemia (AML) in about 30% of patients (5,9,10)**.**

1.2.1.1 Classification

The 2016 edition of the World Health Organization classified MDS into subtypes based on percentage of blasts in the bone marrow and/or peripheral blood, morphological findings, presence of cytopenias, and cytogenetic abnormalities (9,11)**.** The best defined categories are myelodysplastic syndromes with unilineage dysplasia with or without ringed sideroblasts, myelodysplastic syndromes with excess blasts, myelodysplastic syndromes with multilineage dysplasia, and myelodysplastic syndromes with isolated deleted 5q (8) (Table 1).

Table 1 - Diagnostic criteria for MDS according to 2016 WHO classification.

Myelodysplastic Syndromes	Blood findings	Bone-marrow findings
MDS with single lineage dysplasia	One or two cytopenias; no or rare blasts (<1%)	One lineage dysplasia ≥10% of cells in one myeloid lineage; <5% blasts; <15% of erythroid precursors ring sideroblasts
MDS with ring sideroblasts	Anaemia; no blasts	≥15% of erythroid precursors ring sideroblasts; erythroid dysplasia only; <5% blasts
MDS with multilineage dysplasia	Cytopenia(s); no or rare blasts (<1%); no auer rods; $<1x10^9$ cells per L monocytes	Dysplasia in ≥10% of cells in at least two myeloid lineages (neutrophil, erythroid precursors, or megakaryocytes); <5% blasts in marrow; no auer rods; with or without 15% ring sideroblasts
MDS with excess blasts	Cytopenia(s); 5-19% blasts with or without auer rods; $<1x10^9$ cells per L monocytes	Dysplasia in one or several lineages; 5-19% blasts; with or without auer rods

MDS with isolated del(5q)	Anaemia; normal or increased platelet count in most cases; no or rare blasts (<1%)	Normal to increased megakaryocytes with hypolobated nuclei; <5% blasts; isolated del(5q) cytogenetic abnormality; no auer rods
MDS, unclassified	Cytopenias; <1% blasts	Unequivocal dysplasia in <10% of cells in one or more myeloid lineages accompanied by a cytogenetic abnormality is presumptive evidence for diagnosis; <5% blasts

1.2.1.2 Incidence

Myelodysplastic syndromes are the commonest hematological malignancies in the elderly and its incidence keeps rising due to population aging (6,12). The median age at diagnosis of MDS patients is 65-70 years and less than 10% of the patients are younger than 50 years (8). The annual incidence of these disorders is about four cases per 100 000 people, reaching 40-50 per 100 000 after age 70. MDS shows a slight male predominance except for the isolated 5q deletion form in which women predominate. There are also no known ethnic differences in the incidence, but in Asian populations the syndromes tend to occur at an earlier age and be associated with previous disorders (8,13).

1.2.1.3 Etiology

The etiology of MDS is unknown in most patients, although the risk of development of MDS may result from genetic predisposition or hematopoietic stem cell mutations. In a third of pediatric cases is evident an inherited predisposition of the disorder, including in children with Down syndrome, Fanconi anemia, and neurofibromatosis (8). In adults, inherited predisposition to MDS is less common, but has been evident in studies of genetic disorders such as Shwachman-Diamond syndrome and Fanconi anemia and should be investigated in young adults or in families with other cases of MDS, AML, or aplastic anemia (8,14–16).

Myelodysplastic syndromes can result of various factors and depending on the cause of the disorder, MDS can be classified in primary or secondary myelodysplastic syndrome (16,17). Primary or *de novo* myelodysplastic syndromes can result from viral infections, exposure to halogenated organics, metals and petrol compounds, genotoxic chemicals such as benzene, ionizing radiations and cytogenetic abnormalities, such as chromosome

translocations in chromosomes 8 (gain), 5 (loss/deletion), and 7 (loss/deletion). MDS emerging after known exposure to radiotherapy or chemotherapy treatments with alkylating agents or topoisomerase II inhibitors for a primary malignant disease are defined as secondary or therapy-related MDS (t-MDS) (16,17)**.**

Despite being recognized as an important disease for more than 50 years, the molecular pathogenesis of MDS and the molecular basis for its progression to AML remain unclear. Nevertheless, a model of MDS molecular pathogenesis has been proposed whereby a normal hematopoietic stem cell acquires successive genetic abnormalities that ultimately lead to malignant transformation and clonal expansion (6)**.** The genetic predisposition in the etiology of MDS is supported by the observation of defective DNA mismatch repair in patients with t-MDS and significant mitochondrial genomic instability in those with primary MDS (16).

Recently, a large amount of data became available on recurring mutations in MDS. Targeted sequencing of a limited number of genes can detect mutations in 80% to 90% of MDS patients, with the majority of these patients having at least one mutation identified (11,18)**.** The most commonly mutated genes in MDS can be categorized into 4 groups: 1) splicing factors (*SF3B1*, *SRSF2*, *U2AF*); 2) transcription factors (*RUNX1*, *TP53*); 3) epigenetic regulators (*TET2*, *DNMT3A*, *EZH2*, *ASXL1*); and 4) growth signaling factors (*NRAS*, *CBL*, *JAK2*, *SETBP1*) (11,14,18–29)**.**

1.2.1.4 Prognosis

The prognosis of MDS patients is extremely difficult to determine. However, several prognostic systems have been created to predict the clinical outcomes of MDS patients and aid the design and analysis of clinical trials. The first system of prognosis, the International Prognostic Scoring System (IPSS), was created in 1997 (30). This system is based in the combination of the risk scores of three major variables: percentage of blasts in the bone marrow, number of cytopenias, and karyotype. Then, according to their scores patients are stratified into four distinctive risk groups regarding both survival and AML risk evolution: low risk (0), intermediate-1 risk (INT-1; 0.5 to 1.0), intermediate-2 risk (INT-2; 1.5 to 2.0), and high risk (≥2.5) (31). Despite being an important standard for assessing prognosis of primary untreated adult MDS patients, IPSS has since been revised to become more precise predicting the clinical outcomes of MDS patients.

Thus, in 2012 the Revised International Prognostic Scoring System (IPSS-R) was created (32). This system underscores the contribution of chromosomal abnormalities to patients prognosis and has a refined cytogenetic scoring system with specific and new classifications of several less common cytogenetic subsets, splitting the low marrow blast percentage value and depth of cytopenias (31). In this model, patients are stratified in five risk groups [very low risk (0–1,5), low risk (1,5–3), intermediate risk (>3–4,5), high risk (>4,5–6), and very high risk (>6)], according to five main criteria: cytogenetics, bone marrow blast percentage, hemoglobin, platelet count, and absolute neutrophil count levels. Additional variables that provide prognostic information in MDS were also included, such as serum lactate dehydrogenase, ferritin, and β_2-microglobulin, as well as marrow fibrosis and patient comorbidities and performance status.

Besides the two scoring systems already mentioned, there are other scoring systems. One of them, is the World Health Organization (WHO) classification-based Prognostic Scoring System (WPSS), created in 2007 (33). This system was created when additional factors, such as multilineage dysplasia and transfusion dependency, were found to be of additive prognostic value in MDS and were considered a reliable indicator of the severity of the disease, partly reflecting the presence of comorbidities (33,34). The WPSS can be applied at any time during the course of the disease and also classifies patients into five risk groups [very low risk (0), low risk (1), intermediate risk (2), high risk (3-4), and very high risk (5-6)] with different survivals and probabilities of leukemia evolution.

1.2.2 Acute Myeloid Leukemia

Acute myeloid leukemia is a spectrum of progressive malignant diseases characterized by uncontrolled growth of myeloid blood cells in the blood and bone marrow, which leads to a range of clinical problems, including, for example, infection, bleeding, and organ damage (35,36). The pathogenesis of AML is complex, but in general is due to gene duplications and specific gene translocations, point mutations or larger deletions, and a variety of epigenetic changes that lead to under- and overexpression of multiple genes involved in hematopoietic cell growth, differentiation, and self-renewal (35).

1.2.2.1 Classification

The 2016 edition of the World Health Organization defines AML disease entities by focusing on significant cytogenetic and molecular genetic subgroups (11). Many recurring, balanced cytogenetic abnormalities are recognized in AML, and most of those that are not formally recognized by the classification are rare. There are three big groups of AML subtypes: AML with recurrent genetic abnormalities, AML with myelodysplasia-related changes, and not specified AML (Table 2).

Table 2 – 2016 WHO classification of Acute Myeloid Leukemia.

Acute Myeloid Leukemia
AML with recurrent genetic abnormalities
AML with t(8;21)(q22;q22.1);RUNX1-RUNX1T1
AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22);CBFB-MYH11
APL with PML-RARA
AML with t(9;11)(p21.3;q23.3);MLLT3-KMT2A
AML with t(6;9)(p23;q34.1);DEK-NUP214
AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); GATA2, MECOM
AML (megakaryoblastic) with t(1;22)(p13.3;q13.3);RBM15-MKL1
AML with BCR-ABL1
AML with mutated NPM1
AML with biallelic mutations of CEBPA
AML with mutated RUNX1
AML with myelodysplasia-related changes
AML, Not Otherwise Specified (NOS)
AML with minimal differentiation
AML without maturation
AML with maturation
Acute myelomonocytic leukemia
Acute monoblastic/monocytic leukemia
Pure erythroid leukemia
Acute megakaryoblastic leukemia
Acute basophilic leukemia
Acute panmyelosis with myelofibrosis

1.2.2.2 Incidence

Acute myeloid leukemia is the most common acute leukemia in adults, with a median age of 69 years, affecting approximately 15,000 people in the United States each year (35,37). Despite having a lower incidence than many other cancers, the mortality rate of

AML is high, making it the sixth most common cause of cancer-related deaths (35). While rates have improved in the younger age group, the prognosis in older patients remains very poor (38).

1.2.2.3 Etiology

The etiology for most cases of AML is unclear, however the development of the disease has been associated with several risk factors, such as age, genetic disorders, exposure to ionizing radiation, and antecedent hematological disease, as it happens with MDS patients (38). The different subtypes of AML may have distinct mechanisms, suggesting a functional relation between a molecular abnormality or mutation and the causal agent. The most frequent abnormalities found in AML patients include loss or deletion of chromosome 5, 7, Y, and 9, translocations such as t(8;21)(q22;q22); t(15;17)(q222;q11), trisomy 8 and 21, and other abnormalities involving chromosomes 16, 9, and 11. These acquired clonal chromosomal abnormalities are found in 50% to 80% of AML cases (38).

1.3 Aldehyde Dydrogenases Superfamily

The aldehyde dehydrogenase (ALDH) superfamily is a group of $NAD(P)^+$-dependent enzymes that catalyze the irreversible oxidation of a wide variety of endogenous and exogenous aldehydes to their corresponding weak carboxylic acids (39–41)**.** Aldehydes are electrophilic compounds that have a long lifespan and are generated by a wide variety of physiological processes, such as catabolism of amino acids and neurotransmitters like GABA, serotonin, adrenaline, noradrenaline, and dopamine (40–43)**.** Furthermore, there are more than 200 different aldehydes that are produced through some metabolisms: lipid, carbohydrate, vitamins, and steroids (43,44). Along with these endogenous aldehydes, there are also exogenous aldehydes present in a variety of industrial processes, including the production of polyester plastics, smog, cigarette smoke, and motor vehicle exhaust. Other sources of ALDHs are the biotransformation of exogenous aldehyde precursors such as xenobiotics and drugs (e.g. ethanol, cyclophosphamide) (43,44)**.** While many aldehydes play a critical role in physiological processes, most are classified as cytotoxic, mutagenic, and carcinogenic (41–43)**.** These compounds are highly reactive and possess high diffusion capacities in cells. Thus, they can easily form complexes with DNA, proteins, and lipids, altering their function and causing their inactivation. That way the presence of ALDH

enzymes is essential in order to protect the human body from the damaging effects of aldehydes (40).

In the human genome, have been identified 19 ALDH functional genes with distinct chromosomal locations (Table 3) (43,44). ALDHs proteins are classified into families and subfamilies based on the percentage of amino acid identity. Proteins sharing ≥ 40 per cent identity are assigned to a particular family designated by an Arabic numeral, whereas those sharing ≥ 60 per cent identity are classified in the same subfamily designated by a letter (41). Family 1 has 3 subfamilies, family 3 has 2 subfamilies and the other families have 1 subfamily each and every single ALDH gene subfamily is a genetically segregating different cluster of genes located close together in the same chromosomal region (45).

ALDHs have a wide distribution in nature, ranging from bacteria and yeasts to plants and animals (41). In mammalian tissues, ALDHs can be found in the cytoplasm, mitochondria, nucleus, and endoplasmic reticulum. They are constitutively expressed with the highest level in the liver, followed by the kidney, uterus, and brain (39,42,43). These enzymes are involved in a broad spectrum of biosynthesis and metabolism processes and have a wide variety of functions such as catalytic, binding, antioxidant, structural, and regulatory functions (41–43). The ALDH activity is typically a composite of one or more isoforms and depending on enzyme family and subfamily and enzymatic levels. The tissue and organ distribution of the different isoforms can also differ (39,43,46).

Table 3 – ALDH enzymes.

ALDH Isozyme	Subcellular distribution	Cromossomal localization [Reviewd in (47)]	Pathologies associated with altered expression
ALDH1A1	Cytosol	9q21.13	
ALDH1A2	Cytosol	15q21.3	
ALDH1A3	Cytosol	15q26.3	
ALDH1B1	Mitochondria	9q11.1	
ALDH1L1	Cytosol	3q21.3	Neural tube defect (48)
ALDH1L2	Mitochondria	12q23.3	
ALDH2	Mitochondria	12q23.2	Ethanol-induced cancers (49), Cardiovascular diseases (50)
ALDH3A1	Cytosol, nucleus	17p11.2	
ALDH3A2	Microsomes, Peroxisomes	17p11.2	Sjögren-Larsson syndrome (51)
ALDH3B1	Cytosol	11q13	Paranoid schizophrenia

ALDH3B2	Mitochondria	11q13	
ALDH4A1	Mitochondria	1p36	Type II hyperprolinemia (52)
ALDH5A1	Mitochondria	6p22	γ- hydroxybutyric acidúria (53)
ALDH6A1	Mitochondria	14q24.3	Development delay (54)
ALDH7A1	Mitochondria, cytosol, nucleus	5q31	Pyridoxine-depent epilepsy (55)
ALDH8A1	Cytosol	6q23.2	
ALDH9A1	Cytosol	1q23.1	
ALDH16A1	Unknown	19q13.33	Gout and hyperuricemia (56)
ALDH18A1	Mitochondria	10q24.3	Hyperammonemia (57)

1.3.1 ALDH families

The ALDH1 family consist of 3 subfamilies: ALDH1A, ALDH1B, and ALDH1L. The ALDH1A subfamily includes three isoforms, ALDH1A1, ALDH1A2, and ALDH1A3, that are primarily localized in the cytosol of cells from various tissues (45,58)**.** The main biological role of these enzymes is the oxidation of retinal and aliphatic aldehydes (59). These enzymes also play a role in the detoxification of peroxidic aldehydes produced by ultraviolet light absorption, protecting the lens of the eye (45). Among the three conserved cytosolic isozymes of the ALDH1 family, ALDH1A1 is the primary ALDH isoform linked to the elevated activity of ALDH in normal hematopoietic stem cells and cancer stem cell (CSC) populations (60). Therefore, this enzyme plays a vital role as a marker of these populations and has been now commonly used for the isolation of CSCs in multiple tumor settings (44,58,60). The ALDH1A1 enzyme is ubiquitously distributed in adult organs, such as brain, testis, kidney, eye, lens, retina, liver, and lungs and has great affinity for the oxidation of both all-trans-retinal and 9-cis-retinal. Therefore, it is the key enzyme in the oxidation of retinal, the retinol metabolite, to retinoic acid (RA), playing an important role in developing tissues and in the self-renewal, differentiation, and self-protection of stem cells (58)**.**

The second ALDH1 subfamily only contains the isoform ALDH1B1. This isoform is a mitochondrial enzyme expressed in various adult and fetal human tissues including liver, testis, kidney, skeletal muscle, heart, placenta, brain, and lung (44)**.** Despite the scarce knowledge about this enzyme, is known that it displays relatively high affinity for acetaldehyde. Therefore, it is believed that this enzyme plays a major role in acetaldehyde oxidation in the intestine (61)**.** Lastly, the ALDH1L subfamily includes two isoforms

ALDH1L1 and ALDH1L2. The isozyme ALDH1L1 is a folate metabolic enzyme identified as an astrocyte marker (48,62)**.** This cytosolic enzyme is expressed at high levels in the liver, kidney, and pancreas (44). *ALDH1L2* is one of the most recently discovered ALDH genes. ALDH1L2 is a mitochondrial enzyme that has high expression in the liver, kidney, pancreas, heart, and brain (63,64)**.**

Human ALDH2 is a molecule encoded by nuclear DNA, but localized in the mitochondrial matrix (65)**.** ALDH2 is expressed ubiquitously in all tissues but is most abundant in the liver. It can also be found in high amounts in organs that require high mitochondrial oxidative phosphorylation, such as heart and brain. Due to the inactivation of ALDH2 by some reactive aldehyde substrates, it is required a continuously synthesis of this protein (65,66)**.** The ALDH2 family has a high affinity for acetaldehyde and plays an important role in its detoxification, using NAD^+ as a co-factor to the oxidation of acetaldehyde to acetic acid (45,65). It also has been suggested that ALDH2 can contribute to cardiovascular function due to its nitrate reductase activity, since it has the ability to bioactivate nitroglycerin, catalyzing the formation of 1,2-glyceryl dinitrate and nitrite within the mitochondria, which leads to the production of cGMP and vasorelaxation (41,67)**.** Additionally, ALDH2 displays binding capabilities with exogenous compounds, since it has been identified as an acetaminophen binding protein (41)**.**

The ALDH3 family contains enzymes with the capacity to oxidize fatty and peroxidic aldehydes, and enzymes that have noncatalytic functions, including antioxidation functionalities (44,45)**.** These enzymes present a high expression in the cornea, stomach, liver hepatoma cells, and liver microssomes, as well as in some cancer cells (68). To date have been reported four major ALDH3 genes in mammals, *ALDH3A1*, *ALDH3A2*, *ALDH3B1*, and *ALDH3B2* (45)**.** ALDH3A1 is a cytosolic enzyme highly expressed in the stomach, lung, and cornea (68). This enzyme plays a role in cellular oxidative stress related processes and one of the most important functions of this ALDH is the protection of the cornea and the underlying lens against UV-induced oxidative stress (69–72). Furthermore, it is been proposed that ALDH3A1 may scavenge hydroxyl radicals and that it contributes to the antioxidant activity of the cell by producing NADPH, which is linked to the regeneration of reduced glutathione (GSH) from its oxidized form (GSSG) via the glutathione reductase/peroxidase system (41–43)**.** The ALDH3A2 isozyme is expressed ubiquitously displaying activity towards aliphatic aldehydes (70)**.** There are known two

splicing isoforms of the ALDH3A2 enzyme, a major isoform localized in the endoplasmic reticulum and a minor isoform in the peroxisomes. Finally, *ALDH3B1* encodes the human lung, prostate, and kidney cytosolic ALDH, while the *ALDH3B2* encodes the human salivary gland and placental ALDH. Both of these enzymes present a high expression in some tumors (68).

The *ALDH4A1* gene codes for a mitochondrial matrix enzyme highly expressed in the liver, skeletal muscle, and kidney (45). ALDH4A1, also known as pyrroline-5-carboxylate dehydrogenase (P5CDH), is involved in proline degradation and catalyzes the NAD^+-dependent oxidation of γ-glutamate semialdehyde (GSA), the hydrolysis product of pyrroline-5-carboxylate (P5C), to glutamate (73,74).

The *ALDH5A1* gene encodes for succinic semialdehyde dehydrogenase, which is a mitochondrial enzyme that plays a major role in the oxidation of succinic semialdehyde (45,75). This particular enzyme is expressed in the human brain, where it is involved in γ-aminobutyric acid degradation and in the detoxification of lipid peroxidation aldehyde 4-hydroxy-2-nonenal (69).

The ALDH6 family codes for the CoA-dependent methylmalonate semialdehyde dehydrogenase. This enzyme partakes in the degradation of valine and pyrimidines, causing the transformation of malonate and methylmalonate into acetyl- and propionyl-CoA, respectively (45).

ALDH7A1 is an enzyme expressed at high levels in a wide range of tissues, such as the cochlea, eye, ovary, heart, and kidney (44). ALDH7A1 has a primary role in the lysine catabolic pathway by catalyzing the NAD^+-dependent oxidation of α-aminoadipidic-semialdehyde (AASA) to α-aminoadipate (76). Furthermore, ALDH7A1 is also an important biomarker of cancer stem cells and plays a functional role in cancer stem cell-mediated metastasis and cancer drug resistance (76).

The ALDH8A1 enzyme is a cytosolic enzyme with high expression in the liver and kidney (44). Along with the ALDH1 family, the ALDH8A1 isoform is believed to participate in the biosynthesis of RA through conversion of 9-cis-retinal to 9-cis retinoic acid (77,78). Moreover, ALDH8A1 also metabolizes aliphatic aldehydes, including acetaldehyde (44).

The *ALDH9A1* gene encodes γ-trimethylaminobutyraldehyde dehydrogenase, which is the enzyme involved in the metabolism of γ-aminobutyraldehyde and aminoaldehydes derived from polyamines (79). Although these genes are strongly expressed in the early

stages of embryonic brain development, they are poorly expressed in the human adult brain. However, they are present in the liver, kidney, and muscle (45).

The ALDH16 family codes for an enzyme that is widely expressed in a variety of tissues, such as bone marrow, heart, kidney, and lung (45). Although, the physiological significance and function of these enzymes remain unknown (45,80), a study held in Iceland identified a single nucleotide polymorphism in the *ALDH16A1* gene as a risk factor for hyperuricemia and gout (56).

Aldehyde dehydrogenase 18 family, member A1 (*ALDH18A1*) is a gene that encodes for a bifunctional ATP- and NADPH-dependent mitochondrial enzyme, the Δ1-pyrroline-5-carboxylate synthase (P5CS) (45,74). P5CS catalyzes the reduction of L-glutamate to pyrroline-5-carboxylate (P5C), which is a critical step in the *de novo* biosynthesis of proline, ornithine, and arginine (74,81) The ALDH18A1 enzyme is expressed at high levels in the pancreas, ovary, testis, and kidney (45).

All of the ALDH superfamily enzymes play an essential role in the enzymatic detoxification of endogenous and exogenous aldehydes (40). However, it has become clear that most of the ALDHs exhibit multiple and distinct functions within the body due to their varying substrate specificities and gene expression differences (67). Due to the diversity of functions and wide range of tissue distribution of the distinct ALDH families, several mutations and polymorphisms in ALDH genes have been identified. These allelic variants result in distinctive phenotypes including intolerance to alcohol and increased risk of ethanol-induced cancers in most cases (ALDH2) (82) and can cause other severe diseases such as Sjögren-Larsson Syndrome (ALDH3A2) (51), Pyridoxine-dependent epilepsy (ALDH7A1) (55), type II hyperprolinemia (ALDH4A1) (52), 4-hydroxybutyric aciduria, mental retardation and seizures (ALDH5A1) (53), development delay (ALDH6A1) (54), hyperammonemia (ALDH18A1) (57), and neural tube defects (ALDH1L1)(48).

1.4 ALDH in Hematopoietic Stem Cells

Hematopoietic stem cells have a high expression of ALDHs (83). In fact, ALDH enzymes are considered a selectable marker of human HSCs and have been used as a marker to identify and purify these and other stem cells (84,85). Although the specific function of ALDH in HSC remains unknown, some studies suggest that ALDHs promote HSC differentiation (83). In fact, it is known that some ALDH contribute primarily to the

metabolism of retinol into RA (42). Retinoic acid is a low molecular weight (300 Da), lipophilic, rapidly-diffusing signaling molecule essential for growth and developmental processes in the embryo (86)**.** It induces gene transcription and thereby modulates a wide variety of biological processes like cell proliferation, differentiation, cell cycle arrest and apoptosis of different adult cell types (42).

1.4.1 Retinoic Acid Signaling

Retinol (vitamin A) is taken up by cells and is oxidized by retinol dehydrogenases to yield product retinal (Figure 2). Retinal is oxidized by cytoplasmic ALDH enzymes resulting in retinoic acid that diffuses into the nucleus and binds to heterodimers of retinoic acid receptor (RAR) and retinoid X receptor (RXR), activating them. Then, the activated receptor complexes bind to retinoic acid response elements (RAREs), regulatory sequences in target genes, and induce transcription of target genes. Depending on the cellular context, this can lead to differentiation, apoptosis, and/or cell cycle arrest (42)**.**

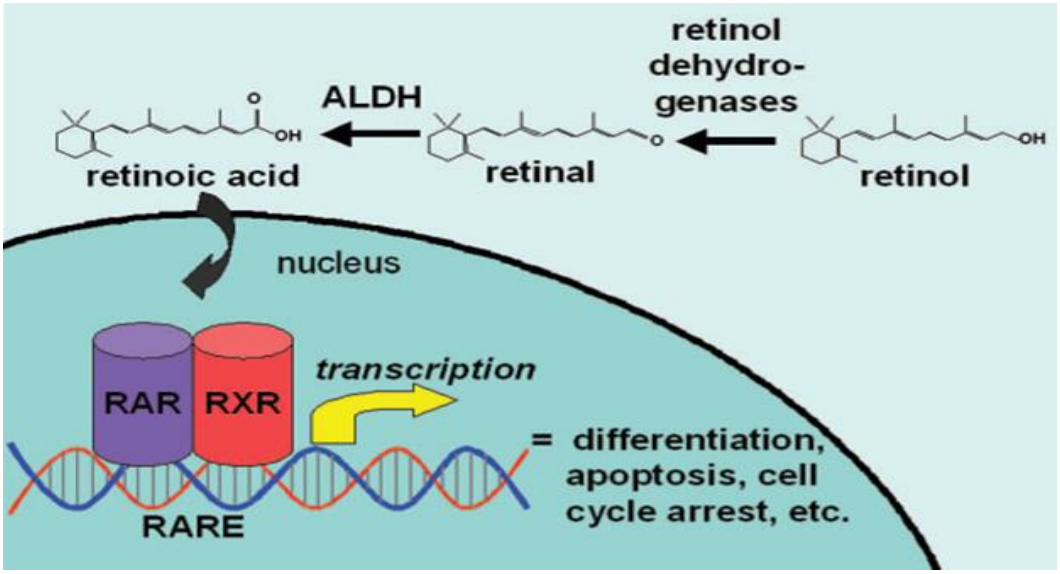

Figure 2 – Retinol Acid Signaling. Retinol is taken up by cells and oxidized by retinol dehydrogenases to yield product retinal. Retinal is oxidized by cytoplasmic ALDH enzymes resulting in retinoic acid. Retinoic acid diffuses into the nucleus and binds to heterodimers of RAR and RXR, activating them. The activated receptor complexes bind to RAREs, inducing the transcription of target genes. Depending on the cellular context, this can lead to differentiation, apoptosis and/or cell cycle arrest. Abbreviations: ALDH, aldehyde dehydrogenases; RAR, retinoic acid receptor; RXR, retinoid X receptor; RARE, retinoic acid response elements. Adapted from (42).

1.4.2 ALDHs and differentiation of HSCs

As ALDHs are highly expressed in HSCs, it is reasonable that the primary function of ALDH in HSCs relates to its production of retinoids (83). As mentioned, the biological actions of retinoids are mediated by the RA and the RX receptors. Through its actions on these receptors, RA induces cellular differentiation. Considering these data, it is plausible that HSC differentiation might depend on retinoid signaling and that ALDHs are essential to the maintenance of the lymphoid and myeloid cell lineages. To date, the enzymes linked to the production of RA are the isoforms ALDH1A1, ALDH1A2, ALDH1A3, and ALDH8A1 (40). In this context, through its role in the production of RA, these ALDH enzymes contribute to HSCs differentiation (83).

1.5 ALDHs in Myelodysplastic Syndromes and Acute Myeloid Leukemia

In recent years, have emerged evidences that suggest tumor progression and metastasis are instigated by a minor cell populations designed by cancer stem cells (CSCs) or tumor-initiating cells (TICs) (60). These cells share characteristics with adult stem cells namely self-renewal, high proliferative potential, and multipotency (87). Thus, as with normal stem cells, several ALDH families are strongly active in CSCs. Due to this high activity, CSCs are being used as a marker for many solid tumors including breast, lung, liver, colon, pancreatic, ovarian, head and neck, prostate, and melanoma (43,88). Since the human ALDH superfamily comprises many isozymes that possess an unique tissue distribution, subcellular localization and substrate specificity, its reasonable that distinct isozymes are expressed in different tumors (46,88). This statement is supported by the reports of elevated expression of ALDH1A1 in melanoma (89) and lung cancer (90), ALDH1A3 in breast cancer (91), ALDH1B1 in colon (92) and pancreatic cancer (93), ALDH3B1 in lung, ovarian, breast, and colon cancer (94), ALDH3A1 in lung cancer (90), and ALDH7A1 in prostate cancer (87).

Due to the critical role of ALDH enzymes in hematopoietic stem cells, it is reasonable to associate these enzymes with the development of hematopoietic disorders, such as MDS and AML (4). Some studies have linked the aldehyde dehydrogenase family with these disorders, since it has been reported an altered expression of some ALDH isoforms in

patients with these diseases (35,36). The pathogenesis of these disorders is characterized by the existence of several mutations in hematopoietic stem cells, so the dysregulation of processes related with differentiation and cellular proliferation of these cells may be associated with the development of these diseases. Some of these biological processes that may be dysregulated in patients with MDS and AML are dependent of ALDH activity, such as the RA synthesis and the oxidation of acetaldehyde. In fact, it is known that, besides its role in hematopoietic stem cells, retinoid signaling pathways also play a significant role in cancer cells. For example, Ginestier *et al.* have reported that ALDH1 regulates breast CSCs by affecting retinoid metabolism (95). In this study, they reported that the blockade of ALDH increases the CSC population and the activation of retinoid signaling reduces the CSC population. Therefore, retinoid signaling modulation may be sufficient to induce the differentiation of breast CSCs (95).

Several studies have found that some stem cells possess higher ALDH activity than differentiated cells**.** In fact, elevated levels of ALDH1A1 are a characteristic feature of hematopoietic stem cells (39)**.** Although, altered expression of ALDH1A1 has been reported in a number of solid tumors and tumor cell lines, the possible roles of this enzyme and the other RA-synthesizing enzymes in hematopoiesis and leukemic transformation remain unexplored (96). However, the dysregulation of RA synthesis due to an increased activity of ALDH enzymes may interfere with processes regulating cellular differentiation and proliferation (Figure 3). One of the target genes that may be involved in the RA-induced proliferation of myeloid leukemia cells is the protooncogene *C-MYC* (86). The dysregulation of this transcription factor is observed in a variety of human malignancies including the leukemias. Thus, it is possible that downregulation of *C-MYC* expression promotes the proliferation of cancer cells, thereby contributing to the tumorigenic phenotype observed in patients with myelodysplastic syndromes (96)**.**

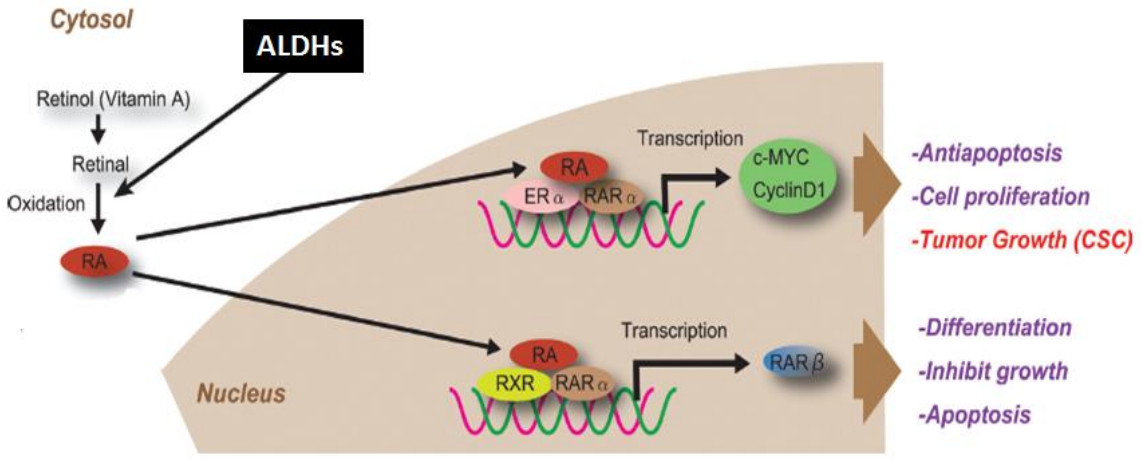

Figure 3 – Regulation and function of ALDH in normal and cancer stem cells. Several ALDHs metabolize RA, thereby regulating the self-renewal and differentiation of hematopoietic and cancer stem cells. Retinal is oxidized to RA by ALDH enzymes. RA binds to RAR and RXR inducing the expression of its downstream target genes, that ultimately lead to differentiation, inhibit growth and apoptosis. In cells that express the ER, RA binds to this receptor inducing the expression of *C-MYC*, a transcription factor that contributes to the proliferation of cancer stem cells. Abbreviations: ALDH, aldehyde dehydrogenase; RA, retinoic acid; RAR, retinoic acid receptor; RXR, retinoid X receptor; RARE, retinoic acid response elements; ER, estrogen receptor. Adapted from (58).

Other biological process that may be altered in MDS and AML patients is the detoxification of acetaldehyde. This aldehyde is the primary reactive metabolite of ethanol and its presence in the organism is due mostly to an excessive alcohol consumption, which can cause a wide range of direct and indirect effects on hematopoiesis, immunity, and cell function (97,98). It has long been known that excess of alcohol use can cause cytopenias, as well as a variety of abnormalities in red cell size and structure. Besides, morphologic analysis of bone marrow samples revealed vacuolization and other abnormalities indicative of the toxic effects resultant from the excessive alcohol consumption. These effects caused by acetaldehyde are in part mediated by protein and DNA adduct formation (Figure 4) (4). These species are more prevalent and, therefore, more damaging in people with reduced DNA adduct repair or reduced ALDH activity (4). This hypothesis is consistent with prior findings that hematopoietic progenitors appear to be more sensitive to ethanol and acetaldehyde than HSCs, possibly because HSCs normally have high levels of ALDH activity that effectively metabolize acetaldehyde to acetate so that DNA and protein adducts do not accumulate. However, people with insufficient ALDHs or DNA repair activity in

HSCs may be more susceptible to accumulating DNA damage from acetaldehyde (4)**.** This hypothesis may seem contradictory, since earlier it is presented that patients with MDS have a higher expression of ALDH. However, there are known polymorphisms in ALDHs responsible for a higher vulnerability to acetaldehyde effects, therefore more susceptible to DNA damage that, ultimately, can lead to mutagenesis and the development of MDS (4).

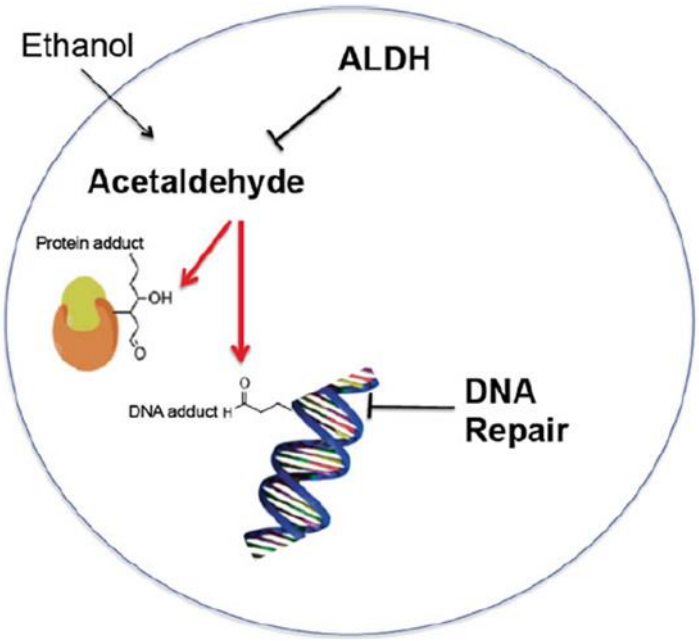

Figure 4 – Model for the possible role of ALDHs in hematopoietic progenitor cells. In this model, deficiencies in ALDH or DNA repair activity in individuals with either polymorphisms or some other abnormality in function will be more susceptible to DNA and proteins damage leading to the development of MDS. These effects may be potentiated in HSCs due to their particularly high levels of ALDH activity that would normally protect these cells from acetaldehyde produced from ethanol. Abbreviations: ALDH, aldehyde dehydrogenases; MDS, myelodysplastic syndromes; HCS, hematopoietic stem cells. Adapted from (4).

The most known polymorphism related with the detoxification of acetaldehyde is the one affecting the ALDH2 isoform. It is caused by a single G to A nucleotide change, which leads to a substitution of glutamate to lysine at amino acid position 487 of the mature protein (E487K). The resultant variant is called ALDH2*2 (E504K) and is the most relevant ALDH2 variant. The single nucleotide E504K polymorphism decreases both the enzyme's stability and activity, which results in a much slower rate of aldehyde metabolism. When this deficiency in ALDH2 expression occurs the detoxification of acetaldehyde is compromised

(65). Recently, the ALDH2*2 genotype has been found to be associated with both an increased risk of sporadic aplastic anemia and more rapid progression of Fanconi anemia, since the carriers of ALDH2*2 enzyme have a lower ALDH2 enzymatic activity (13,65). As MDS, these two diseases are associated with bone marrow failure. Thus, it is possible that this ALDH2 polymorphism may influence MDS development and susceptibility. So, the proliferation of CSCs observed in those patients may result due to an increased mutation rate, resultant of the accumulation of acetaldehyde. This hypothesis may explain why despite the apparent normal or higher ALDH2 expression the development of MDS occurs (4).

Given the critical role of ALDHs in HSCs and in the processes responsible for differentiation and cellular proliferation, it is very likely that these enzymes also have a strong association with the development of hematologic disorders, such as MDS and AML. Despite their specific role in the development of these diseases remain unknown, ALDHs may be used as diagnostic and/or prognostic biomarkers for MDS and AML, due to their altered expression in patients with these two disorders.

Thus, the aim of this study is to evaluate the gene expression of ALDH proteins in patients with MDS and AML in order to identify their potential as a biomarker for diagnosis and/or prognosis of these diseases.

Chapter 2: Methods

2.1 Ethical Statement

The present study was conducted in accordance with the Declaration of Helsinki. The Ethics Committee of Faculty of Medicine of University of Coimbra (Coimbra, Portugal) approved all research procedures. All participants provided their informed consent for participation prior to enrollment.

2.2 Study Population

The present study enrolled a total of 88 individuals: 34 patients with MDS and 20 patients with AML followed in the Hematology Service of Centro Hospitalar e Universitário de Coimbra, EPE (CHUC, EPE), and 34 healthy control individuals. MDS and AML patients were grouped according to the 2016 WHO classification of tumors of hematopoietic and lymphoid tissue. MDS patients were also divided according to the IPSS, IPSS-R, and WPSS risk groups. We collected demographic characteristics for patients and controls, recorded patients' clinical characteristics, such as laboratorial data, and maintained patients' follow-up to collect survival data and evolution to AML.

2.3 RNA Extraction and Quantification

Total RNA was obtained from peripheral blood of patients with MDS and AML and from controls with NZYol reagent (NZYTech) according to the manufacturer's instructions. Briefly, samples were incubated for 5 minutes at room temperature with 1 ml of NZYol. Then, 200 µL of chloroform was added to each sample and the mixture was incubated at room temperature for 3 minutes. The samples were then centrifuged at 12000 gx for 15 minutes to separate the phases. The RNA phase was transferred to another tube, mixed with 500 µL of cold isopropanol, incubated at room temperature for 10 minutes, and centrifuged again for 10 minutes. The supernatant was removed and the pellet was washed with 1 mL of ethanol at 7500 g for 5 minutes. Finally, the ethanol was removed and the pellet dissolved in nuclease-free water. After obtaining the RNA samples, their concentration and degree of purity were quantified by absorption spectrophotometry using NanoDrop® 1000 Spectrophotometer. The degree of purity of the RNA was determined by the ratio 260nm/280nm. The quantification of the RNA samples was determined using their

absorbance at 260 nm, which can be converted to concentration using the Beer-Lambert law. The samples were stored at -80ºC.

2.4 cDNA Synthesis

The RNA extracted from the patients and controls was converted into cDNA with resource to NZY First-Stand cDNA Synthesis Kit, separate oligos, according to manufacturer's protocol. For the cDNA synthesis, total RNA was added to a mixture containing: 0.5 µL of Oligo(Dt)18 primer mix, 0.5 µL of Random hexamer mix, 1 µL of 10x Annealing Buffer, and nuclease-free water. This mixture was incubated at 65ºC for 5 minutes, and then placed on ice. Afterwards, we added 10 µL of NZYRT 2x Master Mix and 2 µL of NZYRT Enzyme Mix and incubated the mixture for 30 min at a temperature of 50 ºC and then at 85ºC for 5 minutes. Finally, we added 1 µL of RNase H (E. coli) and incubate at 37ºC for 20 minutes. The cDNA was stored at -20ºC.

2.5 Analysis of ALDH gene expression levels

The gene expression levels of *ALDH1A1*, *ALDH1A2*, *ALDH1A3*, *ALDH1B1*, *ALDH1L1*, *ALDH1L2*, *ALDH2*, *ALDH3A1*, *ALDH3A2*, *ALDH3B1*, *ALDH3B2*, *ALDH4A1*, *ALDH5A1*, *ALDH7A1*, *ALDH16A1*, and *ALDH18A1* were first analyzed using Reverse Transcriptase Polimerase Chain Reaction (RT-PCR) using the T100TM Thermal Cycler (BIO-RAD). Then, the isoforms ALDH3A2, ALDH3B1, ALDH4A1, and ALDH18A1 were quantified by real time PCR (qPCR). In all reactions were used cell lines as positive and negative controls.

2.5.1 Reverse Transcriptase Polymerase Chain Reaction

The patients and controls cDNA was amplified through conventional PCR technique, using the NZYTaq 2x Green Master Mix Kit (NZYtech) and 0.25 µM of a foward primer and reverse primer (Table 4). The reaction began with an initial denaturation cycle of 2 minutes at 95ºC, followed by 35 cycles of denaturation at 95ºC for 30 seconds, annealing at the temperatures indicated in Table 4, for 30 seconds and extension at 72ºC for 30 seconds. The reaction ended with a final 5 minutes extension cycle at 72ºC. After amplification, the PCR products were analyzed on an agarose gel (NZYtech) at 2%, stained with GreenSafe

(NZYtech) and using a weight marker of 100 bp (NZYDNA Ladder V, NZYtech), and visualized under ultraviolet light.

2.5.2 Real Time Polymerase Chain Reaction

To analyze the ALDHs gene expression, the patients and controls cDNA was amplified through Real Time PCR technique. For this technique, we used the 5x HOT FIREPol® EvaGreen® qPCR Supermix (Solis). To the cDNA was added 0,25 µM of a foward primer and reverse primer (Table 4), 2 µL of 5x HOT FIREPol® EvaGreen® qPCR Supermix, and H2O to complete the 10 µL PCR misture. All samples were used in duplicate and it was included in this analyzes two housekeeping genes, GUS and HPRT. The Real-Time PCR was carried out in a CFX touch (BIO-RAD) in 96-well plates. The thermocycling parameters of the reaction were an initial cycle of 12 minutes at 95ºC, followed by 40 cycles of 15 seconds at 95ºC, 20 seconds at 60ºC, and 20 seconds at 72ºC. The relative experience was calculated with the 2-ΔCT Pfaffl method (99).

Table 4 – Forward and reverse primers of ALDHs.

Gene	ºC	Primer sequence	Length (pb)
ALDH1A1	57	5'-CCTACTCACCGATTTGAAGA-3' 3'-AACATCCTCCTTATCTCCTTC-5'	148
ALDH1A2	59	5'-AGTGTTTTCCAACGTCACTGAT-3' 3'-AGTCTGAGTTATTGGCTCTTTCG-5'	122
ALDH1A3	58	5'-GCCGGGTTCCCTCCA-3' 3'-GTTGAAGAACACTCCCTGAT-5'	261
ALDH1B1	58	5'-CAGAACCCAAGCGTGAT-3' 3'-CTTGCCATTCATTGTTGATGA-5'	193
ALDH1L1	59	5'-TACATGGCATTTATAGACCCG-3' 3'-AGGTTAGACTTCTGTGAGCC-5'	237
ALDH1L2	60	5'-ACCAGTCTGCCGAAGTTTTAC-3' 3'-TTTCTCACCGTCAGTGCTTTT-5'	229
ALDH2	58	5'-ATGGCAAGCCCTATGTCATCT-3' 3'-CCGTGGTACTTATCAGCCCA-5'	94
ALDH3A1	64	5'-CTCTGTGACCCCTCGATCCA-3' 3'-GCATCTTCCCCGTAGAACTCTT-5'	77
ALDH3A2	55	5'-ACTGTGGCTCTAAGAGAGTA-3' 3'-TACCAGGAGCGTCATTAAAA-5'	162
ALDH3B1	58	5'-GCCCTGGAACTATCCGCTG-3' 3'-CGTTCTTGCTAATCTCCGATGG-5'	101

ALDH3B2	56	5’-CTACATATCTCTGCTGTCCG-3’ 3’-GTATAGGGTGGGTAGTGGAT-5’	162
ALDH4A1	63	5’-GACGTGCAGTACCAAGTGTC-3’ 3’-GATCACGGTCTTACCCTGTCC-5’	240
ALDH5A1		5’-ATTTCTTTGAGCCTACCCTG-3’ 3’-TCCTCCTCTGTATCGAACTT-5’	109
ALDH7A1		5’-CCAGTATGCGTGGCTGAAAGA-3’ 3’-CAGGGCAATAGGTCGTAATAACC-5’	104
ALDH16A1	59	5’-CTCAAGGTTCCCCTTAACAC-3’ 3’-TGTTCAGGCTTTAACCACTT-5’	220
ALDH18A1	58	5’-CTGAGTATGGGGACCTGGAA-3’ 3’-GCGGTAACCATCAGAAAAGC-5’	200

2.6 Statistical Analysis

The statistical treatment of the data was performed using the program Statistical Package for Social Sciences (SPSS) version 24. The statistical analysis was carried out by the Kolmogorov-Smirnov Test, Mann-Whitney test, Kruskal-Wallis test, and ROC analysis. A value of $p < 0.05$ was considered significant.

Chapter III: Results

3.1 Characterization of the MDS and AML patients and controls

In this study were enrolled a myelodysplastic syndromes patient group (n=34), an acute myeloid leukemia patient group (n=20), and a healthy control group (n=34) with the characteristics described in Table 5. The MDS group, with a median age of 72 years old, ranging from 49 to 89 years, was composed by 20 females (59%) and 14 males (41%). The AML group, with a median age of 56 years old, ranging from 26 to 92, with 9 females (45%) and 11 males (55%). The group without neoplastic malignancies (healthy controls), was formed by 59% females (n=20) and 41.5% males (n=14) with a median age of 70 years old, ranging from 32 to 88 years. No statistical differences were found between groups in terms of age and gender.

According to the 2016 WHO classification used at the diagnosis, the MDS group included patients with the following subtypes: 4 (12.1%) patients with MDS with single lineage dysplasia (MDS-SLD), 15 (45.5%) with MDS with multilineage dysplasia (MDS-MD), 7 (21.2%) with myelodysplastic syndrome/myelodysplastic-myeloproliferative neoplasm (MDS/MPN), 3 (9.1%) with MDS with excess blasts (MDS-EB), 3 with MDS with ring sideroblasts (MDS-RS) (9.1%), and 1 (3.0%) patient with MDS associated with isolated del(5q) in a total of 33 MDS patients with a known MDS subtype. According to the same classification, the AML group included 7 (35%) patients with AML with minimal differentiation (AML-MD), 5 (25%) with acute myelomonocytic leukemia (AMML), 4 (20%) with acute promyelocytic leukemia with PML-RARA (APL), and 4 (20%) patients with AML with myelodysplasia-related changes (AML-MRD).

The distribution of MDS patients according to IPSS risk system had the following distribution: 13 low risk patients (44.8%), 13 patients with intermediate-1 (44.8%) and 2 with intermediate-2 risk (6.9%), and 1 patient with high risk (3.5%), in a total of 29 MDS patients in which was possible to calculate the IPSS. According to the IPSS-R the patients of the MDS group were distributed in the following risk groups: 6 very low risk patients (20.7%), 17 low risk patients (58.6%), 3 patients with intermediate risk (10.3%), 2 patients with high risk (6.9%), and 1 very high risk patient (3.5%), in a total of 29 MDS patients in which was possible to calculate the IPSS-R. Lastly, the patients were also divided according to the WPSS risk system with the following distribution: 5 very low risk patients (21.7%), 9 low risk patients (39.1%), 6 patients with intermediate risk (26.1%), 2 high risk patients

(8.7%), and 1 patients with very high risk (4.4%), in a total of 23 patients in which was possible to calculate the WPSS.

In 31 patients, we evaluated the existence of cytogenetic abnormalities, which have been grouped by their cytogenetic value according to IPSS. This distribution showed a predominance of abnormalities with good prognostic value [normal karyotype (46XX or 46XY) and del(5q)] in 22 patients (71.0%), 4 patients (12.9%) with intermediate prognostic value [trisomy 8; 46Y/der(X)], 3 patients (9.7%) with poor prognostic value [complex karyotype; del(7q); trisomy 8; del(5)], and 2 patients (6.4%) with normal FISH. Two patients from the total of 33 MDS patients (6.1%) evolved to AML (one of them was classified as MDS-SLD and the other as MDS-MD). The majority of the MDS patients are still alive (n=24; 72.7%) and 9 (27.3%) died. For the AML patient group, 12 (60%) of the 20 patients have died and 8 (40%) are still alive.

Table 5 – Demographic and clinical characteristics of MDS and AML patients and control individuals.

Characteristics	**MDS (n=34)**		**AML (n=20)**		**Controls (n=34)**	
	n	%	n	%	n	%
Demographic Data						
Gender						
Female	20	59	9	45	20	59
Male	14	41	11	55	15	41
Age (years)						
Median age	72		56		70	
Range	49-89		26-92		32-88	
Clinical Data						
AML classification (n=20)						
AML-MD			7	35		
APL			5	25		
AMML			4	20		
AML-MRC			4	20		
MDS classification (n=33)						
MDS-SLD	4	12.1				

MDS-MD	15	45.5
MDS/NMP	7	21.2
MDS-EB	3	9.1
MDS-RS	3	9.1
MDS-del(5q)	1	3.0
IPSS (n=29)		
Low	13	44.8
Intermediate-1	13	44.8
Intermediate-2	2	6.9
High	1	3.5
IPSS-R (n=29)		
Very Low	6	20.7
Low	17	58.6
Intermediate	3	10.3
High	2	6.9
Very High	1	3.5
WPSS (n=23)		
Very Low	5	21.7
Low	9	39.1
Intermediate	6	26.1
High	2	8.7
Very High	1	4.4
Cytogenetic Abnormalities (n=31)		
Good prognostic	22	71.0
Intermediate prognostic	4	12.9
Poor prognostic	3	9.7
Normal FISH	2	6.4

MDS, myelodysplastic syndrome; AML, acute myeloid leukemia; AML-MD, AML with minimal differentiation; APL, acute promyelocytic leukemia with PML-RARA; AMML, acute myelomonocytic leukemia; AML-MRC, AML with myelodysplasia-related changes; MDS-SLD, MDS with single lineage dysplasia; MDS-MD, MDS with multilineage dysplasia; MPN, myeloproliferative neoplasm; MDS-EB, MDS with excess blasts; MDS-RS, MDS with ring sideroblasts; MDS-del(5q), MDS with long-arm deletion of chromosome 5.

3.2 Evaluation of the expression levels of ALDHs by Reverse Transcriptase PCR

Initially we performed RT-PCR in the *ALDH1A1*, *ALDH1A2*, *ALDH1A3*, *ALDH1B1*, *ALDH1L1*, *ALDH1L2*, *ALDH2*, *ALDH3A1*, *ALDH3A2*, *ALDH3B1*, *ALDH3B2*, *ALDH4A1*, *ALDH5A1*, *ALDH7A1*, *ALDH16A1*, and *ALDH18A1* isoforms to see which ones appeared to have different levels of expression between the groups in study (Figure 5). This step allowed the selection of the ALDHs genes to quantify by real time PCR. The isoforms that showed different expression profiles between the groups in this first analysis were the *ALDH1A1*, *ALDH1A2*, *ALDH1B1*, *ALDH2*, *ALDH3A2*, *ALDH3B1*, *ALDH4A1*, *ALDH7A1*, *ALDH16A1*, and *ALDH18A1* isoforms.

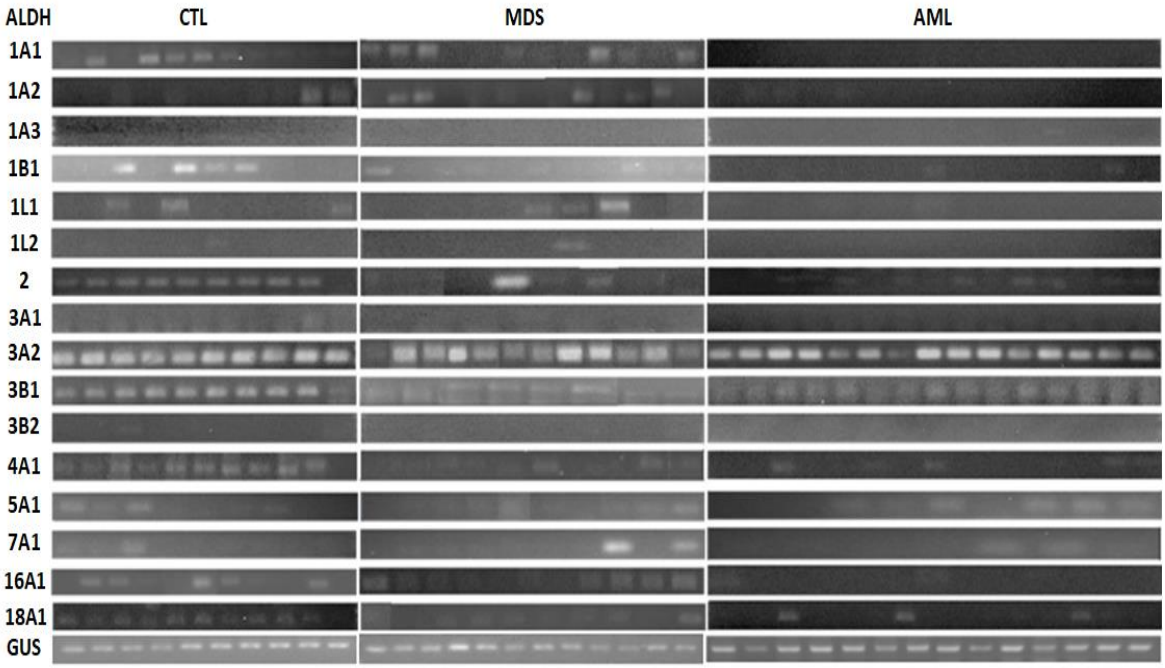

Figure 5 – ALDH gene expression profile of healthy controls, MDS patients, and AML patients. ALDH isoform expressions were analyzed by RT-PCR. *ALDH1A1*, *ALDH1A2*, *ALDH1B1*, ALDH2, *ALDH3A2*, *ALDH3B1*, *ALDH4A1*, *ALDH7A1*, *ALDH16A1*, and *ALDH18A1* exhibited differential gene expression between healthy controls and MDS and AML patients. Abbreviation: CTL, healthy control; MDS, myelodysplastic syndrome patient; AML, acute myeloid leukemia patient.

3.3 Evaluation of ALDH3A2, ALDH3B1, ALDH4A1, and ALDH18A1 gene expression levels in MDS and AML patients and controls

The expression levels of *ALDH3A2*, *ALDH3B1*, *ALDH4A1*, and *ALDH18A1* genes were compared between MDS and AML patients and control individuals (Figure 6). For the *ALDH3A2* gene we observed that MDS patients (median 1.9251; interquartile range 1.28) and AML patients (median 1.5096; interquartile range 0.99) had higher expression levels than controls (median 0.4624; interquartile range 1.53), with statistical differences between MDS and controls (p=0.000618) as well as AML and controls (p=0.008). No statistical differences were found between MDS and AML patients (p=0.716). On the other hand, MDS patients (median 1.6445; interquartile range 1.39) had higher *ALDH3B1* expression levels in comparison with controls (median 0.3521; interquartile range 0.51; p=5.9942E-07) and with AML patients (median 0.4541; interquartile range 0.47; p=0.000314). However, no statistical differences were found between controls and AML patients (p=0.267). *ALDH4A1* gene expression levels were significantly higher in MDS patients (median 0.1841; interquartile range 0.47) when compared to controls (median 0.0388; interquartile range 0.12; p=0.011349). Moreover, *ALDH4A1* expression levels were also higher in AML patients (median 0.1635; interquartile range 0.78) in comparison with controls (p=0.124), however without statistical significance. Lastly, the *ALDH18A1* gene presented similar expression levels in AML patients (median 0.0644; interquartile range 0.22) and controls (median 0.0339 interquartile range 0.07; p=0.425). However, this isoform was not expressed in any MDS patients and in some AML patients.

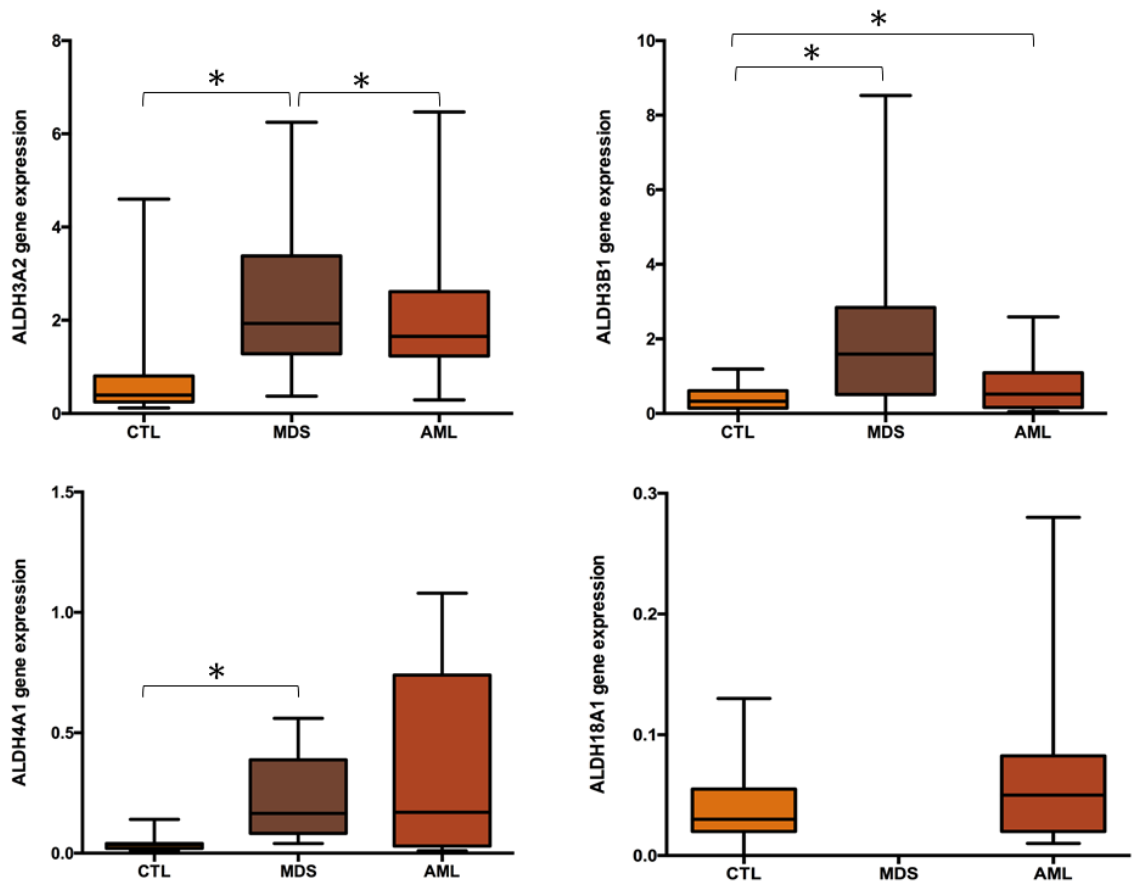

Figure 6 – Analysis of the gene expression of *ALDH3A2*, *ALDH3B1*, *ALDH4A1*, and *ALDH18A1* in healthy controls, MDs patients, and AML patients. MDS and AML patients had higher expression of *ALDH3A2* than controls, with statistical differences between MDS and controls (p=0.000618) as well as AML and controls (p=0.008). MDS patients also had higher *ALDH3B1* expression levels in comparison with controls (p=5.9942E-07) and with AML patients (p=0.000314). *ALDH4A1* gene expression levels were significantly higher in MDS patients when compared to controls (p=0.011349). Lastly, the *ALDH18A1* gene presented similar expression levels in AML patients and controls. However, this isoform was not expressed in any MDS patients and in some AML patients.

3.4 Evaluation of ALDH expression levels according to MDS and AML subtypes of the World Health Organization classification

Next, we analyzed the expression levels of the *ALDH3A2*, *ALDH3B1*, *ALDH4A1*, and *ALDH18A1* genes according to the MDS (Figure 7) and AML subgroups. The *ALDH3A2* gene expression levels were statistical different between the expression levels of the patients of different subgroups (p=0.016). The MDS EB was the MDS subgroup with higher expression levels and MDS RS had the lower expression levels. It was also showed statistically significant differences in *ALDH3A2* expression between MDS RS (median 1.0717) and MDS MD (median 2.3157; interquartile range 2.18; p=0.016), MDS RS and

MDS EB (median 3.3437; interquartile range 1.83; *p*=0.006), MDS/NMP (median 1.3376; interquartile range 1.53) and MDS MD (*p*=0.019), and between MDS/NMP and MDS EB (*p*=0.008). All the other relations between each two MDS subgroups do not show significant differences. The *ALDH3B1* gene expression levels also showed statistical differences between the different subgroups (*p*=0.032). This isoform was significantly higher in MDS SLD patients (median 18.5071; interquartile range 11.81) when compared to MSD RS (median 1.0134; *p*=0.047), MDS/NMP (median 1.1646; interquartile range 1.28; *p*=0.031), and MDS MD (median 1.8195; interquartile range 2.65; *p*=0.016). Lastly, the MDS subgroups did not show differences in *ALDH4A1* gene expression levels (*p*=0.336).

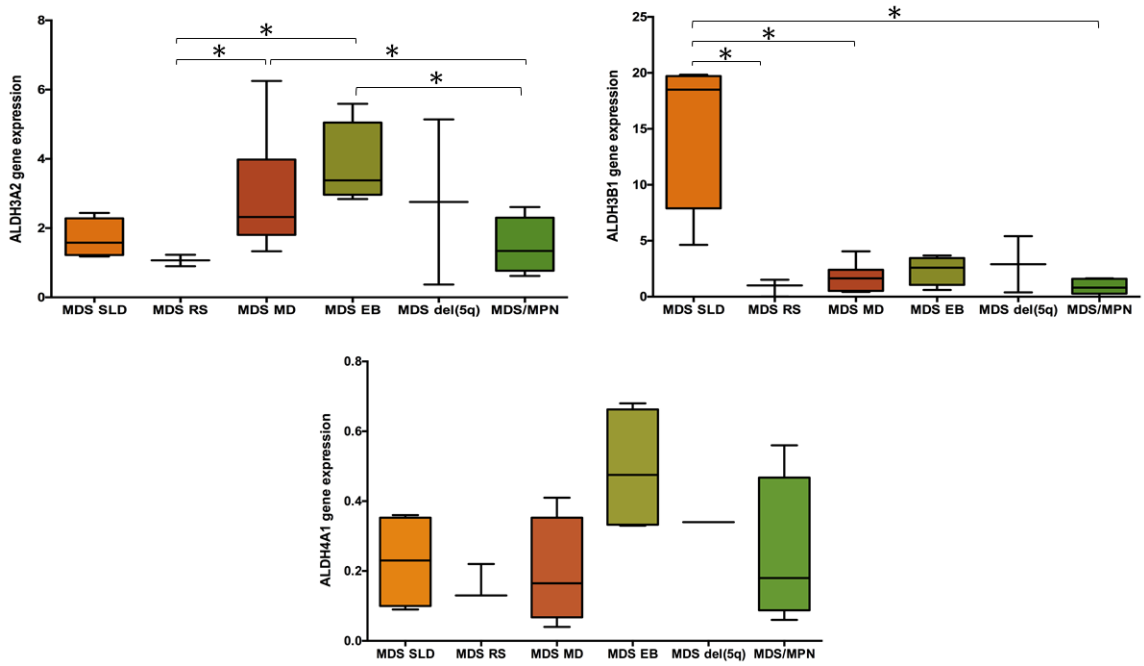

Figure 7 – Analysis of *ALDH3A2*, *ALDH3B1*, and *ALDH4A1* in MDS patients according to the 2016 WHO classification. *ALDH3A2* gene expression levels were statistical different between the subgroups: MDS RS and MDS MD (p=0.016), MDS RS and MDS EB (p=0.006), MDS/NMP and MDS MD (p=0.019), and MDS/NMP and MDS EB (p=0.008). *ALDH3B1* gene expression levels were statistically different between the subgroups: MDS SLD and MSD RS (p=0.047), MDS SLD and MDS/NMP (p=0.031), and MDS SLD and MDS MD (p=0.016). Lastly, the MDS subgroups did not show differences in *ALDH4A1* gene expression levels (p=0.336).

For the AML patients, the gene expression levels of the four ALDHs analyzed showed no statistically differences between the 2016 WHO classification subgroups (Figure 8).

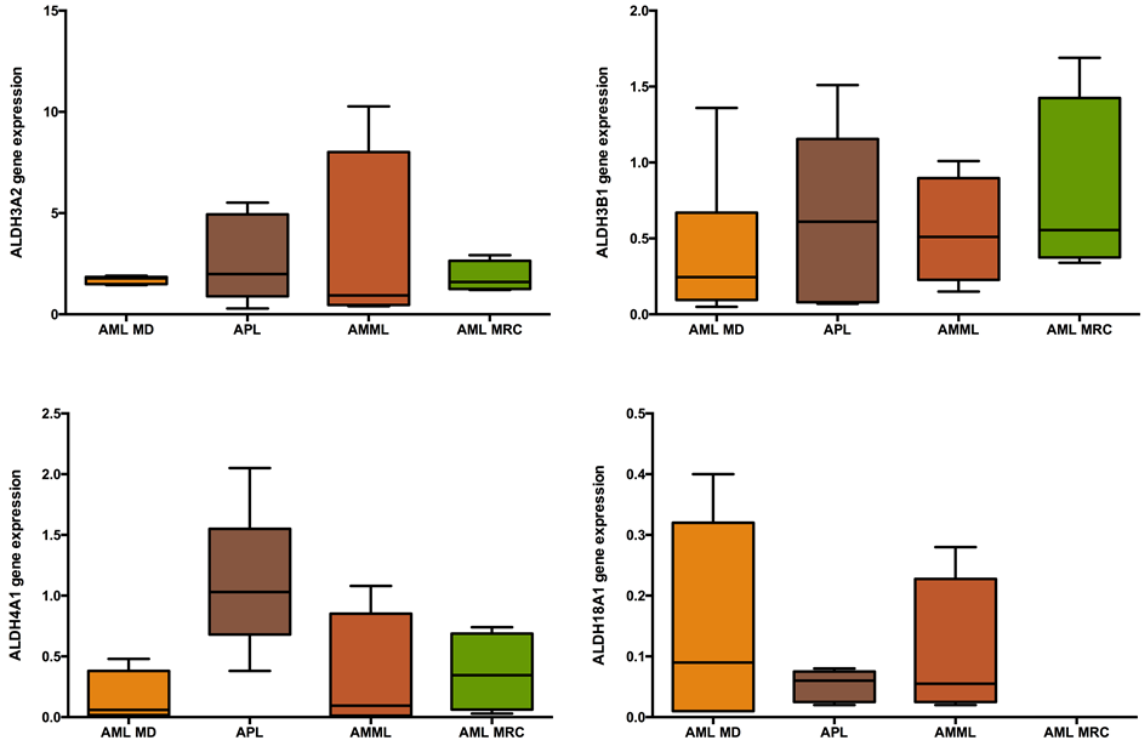

Figure 8 – Analysis of *ALDH3A2*, *ALDH3B1*, *ALDH4A1*, and *ALDH18A1* in AML patients according to the 2016 WHO classification. The four ALDHs analyzed showed no statistically differences between the subgroups of AML.

3.5 Analysis of ALDH expression levels according to IPSS, IPSS-R and WPSS

We analyzed the expression levels of the *ALDH3A2*, *ALDH3B1*, and *ALDH4A1* genes according to the IPSS risk groups. To determine the contribution of *ALDH3A2*, *ALDH3B1*, and *ALDH4A1* in MDS prognosis, patients were grouped according to their IPSS risk (Figure 9), but no statistically significant differences were observed between the expression levels of these genes and the patients of the four IPSS subgroups (*ALDH3A2* p=0,547; *ALDH3B1* p=0,212; *ALDH4A1* p=0,288).

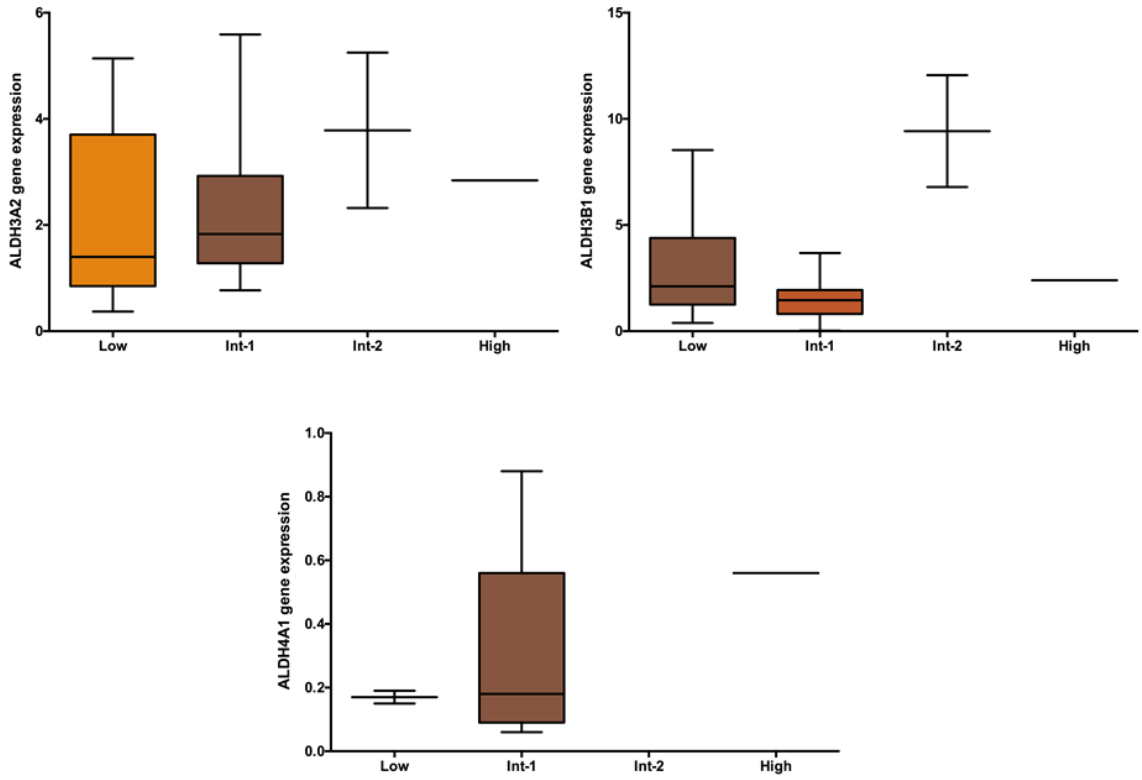

Figure 9 – Analysis of *ALDH3A2*, *ALDH3B1*, and *ALDH4A1* according to IPSS risk groups. The three ALDHs analyzed showed no statistical differences between the different group risks of IPSS.

Next, we evaluated the ALDH expression levels according to IPSS-R score (Figure 10). This system divides the patients in five risk groups: very low risk, low risk, intermediate risk, high risk, and very high risk. No significant differences were observed between these groups for all the ALDH genes (*ALDH3A2* p=0,135; *ALDH3B1* p=0,725; *ALDH4A1* p=0,118).

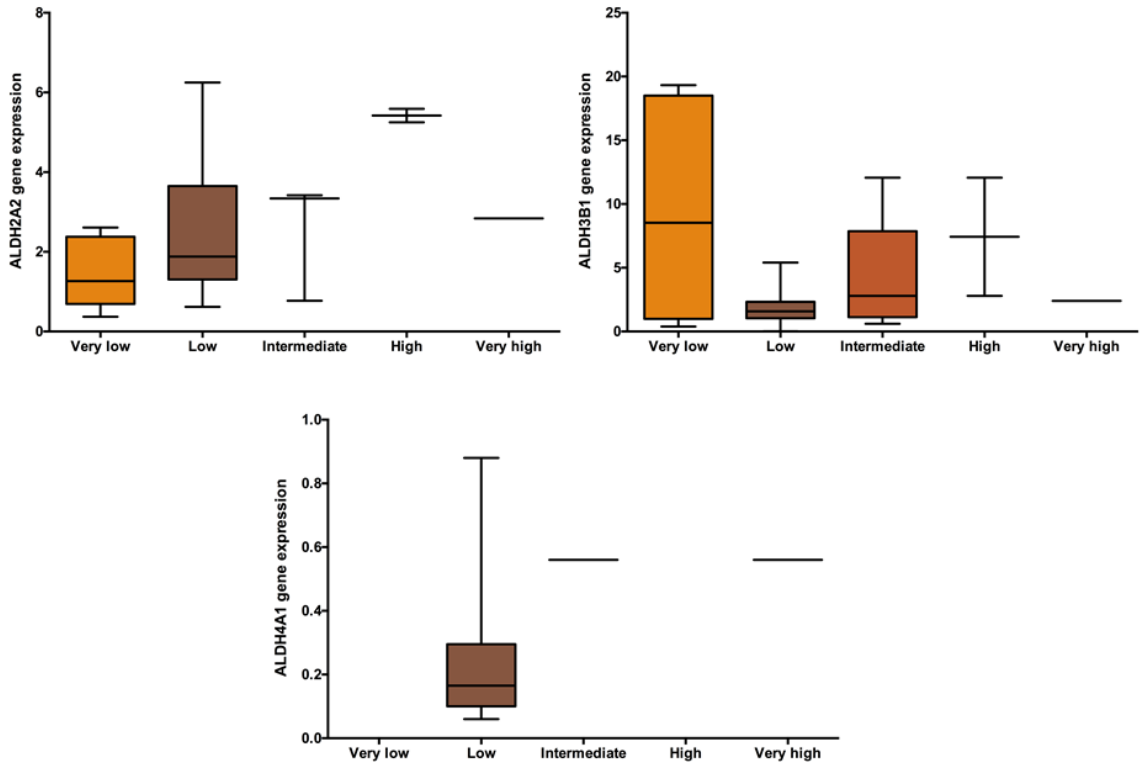

Figure 10 – Analysis of *ALDH3A2*, *ALDH3B1*, and *ALDH4A1* according to IPSS-R risk groups. The three ALDHs analyzed showed no statistical differences between the different group risks of IPSS-R.

Finally, we analyzed the expression levels of the *ALDH3A2*, *ALDH3B1*, and *ALDH4A1* genes according to the WPSS risk groups (Figure 11). According to this score system we observed that the *ALDH3A2* (p=0.761) and *ALDH4A1* (p=0.300) showed no statistically significant differences between their expression levels and the patients of the WPSS subgroups. However, MDS patients with very low risk showed higher expression levels of *ALDH3B1* in comparison with the other risk groups (p=0.035). Moreover, MDS patients with very low risk had significant higher levels of *ALDH3B1* than low risk patients (p=0.019) and intermediate risk patients (p=0.010), and intermediate risk patients showed lower expression levels of this isoform when compared to high risk patients (p=0.035).

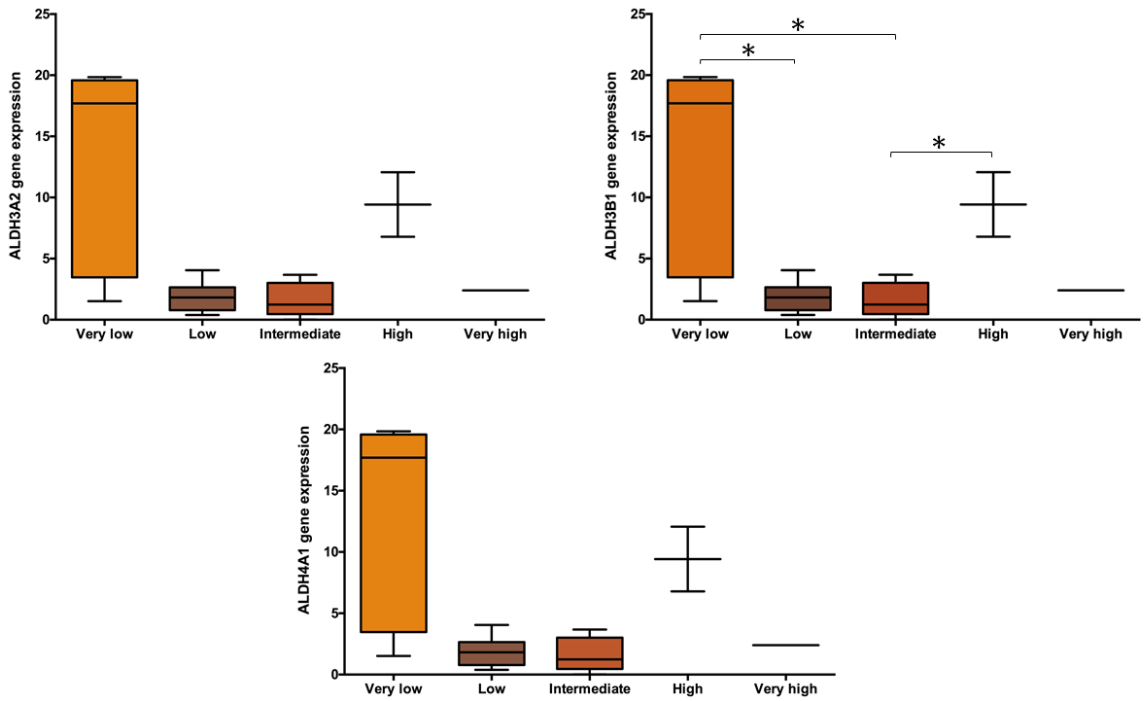

Figure 11 – Analysis of *ALDH3A2*, *ALDH3B1*, and *ALDH4A1* according to WPSS risk groups. According to this score system, the *ALDH3A2* and *ALDH4A1* showed no statistically significant differences between their expression levels and the patients of the WPSS subgroups. *ALDH3B1* has a higher expression in MDS patients with very low risk than low risk patients (p=0.019) and intermediate risk patients (p=0.010), and intermediate risk patients showed lower expression levels of this isoform when compared to high risk patients (p=0.035).

3.6 Evaluation of ALDH3A2, ALDH3B1, ALDH4A1, and ALDH18A1 as MDS and AML diagnostic biomarkers

To evaluate if the *ALDH3A2*, *ALDH3B1*, and *ALDH4A1* isoforms could to be used as MDS diagnostic markers, we determine the capacity of the gene expression levels of these ALDHs to discriminate MDS patients from controls in peripheral blood by ROC analysis. As we can observe in the Table 6, the ROC curve analysis showed that *ALDH3A2* (p=0.001), *ALDH3B1* (p=0.000001), and *ALDH4A1* (p=0.012) were able to discriminate MDS patients from controls. However, the gene that exhibits a higher discriminatory capacity was the *ALDH3B1*, with an area under the curve (AUC) of 0.848 [95% confidence interval (CI): 0.753-0.944]. *ALDH3B1* levels higher than 1.0866 were the optimal cut off value for differentiation of MDS patients from controls. This cut off presented high sensitivity (91%) and specificity (70%), as well as a good power to rule in patients [positive predictive value (PPV): 77%] and to rule out controls [negative predictive value (NPV): 74%]. Since

ALDH18A1 was not expressed in any MDS patients we were unable to perform its statistical analysis. However, this isoform may be the more accurate diagnostic biomarker of MDS.

Table 6 - Performance of ALDH3A2, ALDH3B1, and ALDH4A1 to discriminate MDS patients from controls.

AUC			Cut Off				
ALDH	**95% IC**	***p***	**Value**	**Sen (%)**	**Spe (%)**	**PPV (%)**	**PNV (%)**
3A2	0.742 (0.609-0.876)	0.001	0.7244	66	94	73	79
3B1	0.848 (0.753-0.944)	0.000001	1.0866	91	70	77	74
4A1	0.767 (0.622-0.911)	0.012	0.0547	70	90	50	96

ALDH, aldehyde dehydrogenase; AUC, area under the curve; 95% CI, 95% confidence interval; Sen, sensitivity; Spe, specificity; PPV, positive predictive number; PNV, negative predictive number.

We also determine the capacity of ALDH gene expression levels to discriminate AML patients from controls. According to the ROC curve analysis (Table 7), *ALDH3A2* was the only isoform with value for discriminating AML patients from controls, with an AUC of 0.716 (95% CI: 0.575-0.856; p=0.008). For this gene, the optimal expression levels cut off was higher than 0,9194, presenting good sensitivity (69%) and specificity (86%), as well as a good power to rule in [positive predictive value (PPV): 64%] and to rule out [negative predictive value (NPV): 92%]. The expression levels of the other ALDHs evaluated (*ALDH3B1*, *ALDH4A1*, and *ALDH18A1*) were not able to discriminate AML patients from controls. However, as observed in MDS patients, AML-MRC patients did not express *ALDH18A1* indicating that this isoform may be a diagnostic biomarker for AML-MRC and being correlated with myelodysplasia.

Table 7 - Performance of ALDH3A2, ALDH3B1, ALDH4A1, ALDH18A1 to discriminate AML patients from controls.

	AUC			Cut Off			
ALDH	**95% IC**	***p***	**Value**	**Sen (%)**	**Spe (%)**	**VPP (%)**	**VPN (%)**
3A2	0.716 (0.575–0.856)	0.008	0.9194	69	86	64	92
3B1	0.429 (0.438-0.817)	0.455	nc	nc	nc	nc	nc
4A1	0.599 (0.425-0.774)	0.298	nc	nc	nc	nc	nc

ALDH, aldehyde dehydrogenase; AUC, area under the curve; 95% CI, 95% confidence interval; Sen, sensitivity; Spe, specificity; PPV, positive predictive number; PNV, negative predictive number; nc, not calculated.

Chapter IV: Discussion

The clinical importance of the ALDH superfamily is evidenced by the multiple diseases associated with ALDH dysfunction. Besides its role in the pathogenesis of several cancers, ALDHs have also been associated with the development of hematopoietic disorders, such as MDS and AML, due to the critical role of ALDH enzymes in hematopoietic stem cells (4). Although the mechanisms and isoforms that contribute to the development of these diseases remain unknown, some studies have linked the aldehyde dehydrogenase family with these disorders, since it has been reported an altered expression of some ALDHs isoforms in patients with these diseases (35,36). However, ALDHs are enzymes with different distributions and functions in the organism, so not all would have implications in the pathogenesis of MDS and AML. In order to evaluate the differences in the expression profiles of the different ALDH enzymes in MDS and AML patients, and healthy controls we analyzed the gene expression levels of *ALDH1A1*, *ALDH1A2*, *ALDH1A3*, *ALDH1B1*, *ALDH3A1*, *ALDH3A2*, *ALDH3B1*, *ALDH3B2*, *ALDH4A1*, *ALDH5A1*, *ALDH7A1*, *ALDH16A1*, and *ALDH18A1*. The initial qualitative expression analysis revealed that some ALDH isoforms, such as *ALDH1A3*, *ALDH1L1*, *ALDH1L2*, *ALDH3A1*, *ALDH3B2*, *ALDH5A1*, did not have significant differences in the expression levels between the three groups in study.

In this context, the expression levels of differentiated expression of some ALDHs (*ALDH3A2*, *ALDH3B1*, *ALDH4A1*, and *ALDH18A1*) were quantified by qPCR. Our results showed an increased expression of *ALDH3A2* in the MDS and AML patient groups when compared to healthy controls. In MDS patients an increase in the *ALDH3B1* and *ALDH4A1* genes were also observed comparatively with the control group. *ALDH3B1* had also a higher expression in MDS patients than in the AML patients. Lastly, the *ALDH18A1* expression levels were higher in patients with AML than in controls, except in patients with AML-MRC that did not expressed this ALDH isoform. Also, none of the patients with MDS showed expression of *ALDH18A1*. According to the literature, the loss of ALDHs in HSCs may promote malignant transformation of these cells (36). A study using primitive murine hematopoietic cells suggested that the lack of *ALDH1A1* and *ALDH3A1* in these cells may increase their propensity for leukemic transformation. In our study, we found that the expression of these enzymes can be very heterogeneous. To the best of our knowledge, only one study analyzed the expression levels of some ALDHs in hematological malignancies. In this study, *Yin* (2011) found that *ALDH1A3* levels were lower in bone marrow mononuclear

cells (BMMCs) from AML patients comparatively to healthy donors, while *ALDH3A1*, *ALDH5A1*, and *ALDH8A1* were higher in BMMCs from AML patients in comparison to healthy donors (99). Here we found increased levels of *ALDH3A2*, *ALDH3B1* and *ALDH4A1* in peripheral blood mononuclear cells (PBMCs) from AML patients in comparison to controls. These differences may be due to sampling, since we analyzed PBMCs and the other study analyzed BMMCs. Moreover, the loss of *ALDH18A1* expression in MDS and AML-MRC found in the present study suggests its involvement in the myelodysplasia development.

The correlation of ALDHs with normal and cancer stem cells is reported since the 1980s, with reports describing an overexpression of ALDHs in leukemic cells (100). It was also discovered that the expression of ALDHs was a characteristic of healthy progenitor hematopoietic cells but was gradually lost during the maturation process. Since then, ALDHs are often used as markers for isolating hematopoietic progenitor populations and, more recently, also used as CSC markers in leukemias and solid tumors. Until now, ALDHs used as CSC markers have been investigated in a broad range of different cancers and it were found to be promising markers for the discrimination of sub-populations with stem-like characteristics. However, there is still a long way to identifying the specific ALDHs isoforms responsible for different types of cancer (100). To assess the accuracy of ALDH expression levels as possible diagnostic biomarkers for MDS and AML, it was performed a ROC curve analysis. In the AML patient group, the ROC analysis only showed that *ALDH3A2* expression levels may be a possible diagnostic biomarker for AML. Additionally, the ROC curve also showed that *ALDH3A2*, *ALDH3B1*, and *ALDH4A1* expression levels may be good diagnostic biomarkers. Among them, the gene that exhibits a higher discriminatory capacity and could be used as a positive diagnostic biomarker is the *ALDH3B1*. However, according to the results it was also found a gene that could be used as a negative diagnostic biomarker for MDS and AML-MRC, the *ALDH18A1*. Since none of the MDS and AML-MRC patients showed any expression of this enzyme, it could be used to distinguish healthy individuals from individuals these pathologies and help in their diagnosis. For more than fifty years that it had been suggested that mitochondrial defects lead to an altered metabolism that is specific of cancer cells (101). In fact, mitochondrial defects may be another factor involved in the pathogenesis of MDS and AML. There is evidence of an increased apoptosis and reduced mitochondrial membrane potential in bone marrow cells from MDS patients (102). Besides

that, in MDS patients' mitochondria often show ultra-structural abnormalities, such as pathological iron accumulation. Moreover, mitochondrial generated ATP and mitochondrial localized metabolic pathways have been shown to play an important role in the progression of AML (101). A study conducted in patients with MDS and AML-MRC found that the expression of mitochondrial-encoded genes was significantly reduced in these patients when compared to normal controls (103). Since, ALDH18A1 is a mitochondrial enzyme, it is possible that the lack of expression of ALDH18A*1* in patients with MDS and AML-MRC may be correlated with the mitochondrial dysfunction observed in these diseases.

As previously mentioned, there are different subtypes of MDS and AML (11). Knowing that, we analyzed the expression levels of *ALDH3A2*, *ALDH3B1*, *ALDH4A1*, and *ALDH18A1* in patients with different MDS and AML subtypes to assess if there were any differences in the expression profiles of these enzymes according to subtype of the diseases. The *ALDH3A2* enzyme has a very heterogeneous expression profile among the different subtypes of MDS, but the patients with a higher expression of this enzyme are the ones with MDS EB and the patients with MDS RS have the lowest expression of *ALDH3A2*. Moreover, the *ALDH3B1* enzyme is mostly expressed in patients with MDS SLD. In AML subgroups, we did not see differences between them, except in AML-MRC patients that, like MDS patients, did not expressed *ALDH18A1*.

Lastly, it was assessed the contribution of *ALDH3A2*, *ALDH3B1*, and *ALDH4A1* in MDS prognosis, to evaluate if the expression levels of these enzymes could improve the MDS scoring risk systems. Patients were grouped according to their IPSS, IPSS-R and WPSS risk groups. For the first two risk scoring systems, none of the enzymes showed statistically significant differences between the different risk groups. However, according to WPSS risk groups, the expression levels of the *ALDH3B1* enzyme were higher in low risk patients when compared to other risk groups.

The present study shows some limitations. One of these limitations is associated with the sampling. Besides having a lower number of individuals that participated in the study, the MDS patient group was predominantly constituted by patients with MDS MD and low risk patients, making it impossible to analyze properly correlations between the ALDHs and all the MDS subtypes or high-risk patients. The AML group present the same limitations. Moreover, not all ALDH isoforms were studied. In this context, prospective studies enrolling a significant number of MDS and AML patients will be needed to confirm our results.

Chapter V: Conclusion

In this study, we suggest that ALDHs could be used as MDS and AML diagnostic biomarkers. The present findings indicate that the enzymes *ALDH1A3*, *ALDH1L1*, *ALDH1L2*, *ALDH3A1*, *ALDH3B2*, and *ALDH5A1* do not have significant differences in the expression levels between healthy controls, MDS patients, and AML patients. However, we found that the expression levels of *ALDH3B1* can be a positive diagnostic biomarker and that *ALDH3A2* may be a positive diagnostic biomarker in the diagnosis of AML. Moreover, we also found that *ALDH18A1* could be a very good negative diagnostic biomarker of MDS and AML-MRC, since none of these patients showed expression of this isoform. However, further studies are needed to prove the potential of these enzymes as diagnostic biomarkers.

Chapter VI: References

1. Seita J, Weissman IL. Hematopoietic stem cell: self-renewal versus differentiation. Wiley Interdiscip Rev Syst Biol Med. 2010;2(6):640–53.

2. Hu D, Shilatifard A. Epigenetics of hematopoiesis and hematological malignancies. Genes Dev. 2016;30(18):2021–41.

3. Ogawa M. Differentiation and proliferation of hematopoietic stem cells. Blood. 1993;81(11):2844–53.

4. Smith C, Gasparetto M, Jordan C, Pollyea DA, Vasiliou V. The Effects of Alcohol and Aldehyde Dehydrogenases on Disorders of Hematopoiesis. In: Advances in experimental medicine and biology. 2015. p. 349–59.

5. Goldberg SL, Chen E, Corral M, Guo A, Mody-Patel N, Pecora AL, et al. Incidence and Clinical Complications of Myelodysplastic Syndromes Among United States Medicare Beneficiaries. J Clin Oncol. 2010;28(17):2847–52.

6. Look AT. Molecular Pathogenesis of MDS. Hematology. 2005;2005(1):156–60.

7. Prebet T, Zeidan A. Trends in Clinical Investigation for Myelodysplastic Syndromes. Clin Lymphoma Myeloma Leuk. 2016;16:S57–63.

8. Adès L, Itzykson R, Fenaux P. Myelodysplastic syndromes. Lancet. 2014;383(9936):2239–52.

9. Canaani J, Nagler A. Established and emerging targeted therapies in the myelodysplastic syndromes. Expert Rev Hematol. 2016;9(10):997–1005.

10. Poynter JN, Richardson M, Roesler M, Blair CK, Hirsch B, Nguyen P, et al. Chemical exposures and risk of acute myeloid leukemia and myelodysplastic syndromes in a population-based study. Int J cancer. 2016;140(1):23–33.

11. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood. 2016;127(20):2391–405.

12. Nomdedeu M, Pereira A, Ramos F, Valcárcel D, Costa D, Arnan M, et al. Excess Mortality in the Myelodysplastic Syndromes. Am J Hematol. 2016;92(2):149–154.

13. Hira A, Yabe H, Yoshida K, Okuno Y, Shiraishi Y, Chiba K, et al. Variant ALDH2 is associated with accelerated progression of bone marrow failure in Japanese Fanconi anemia patients. Blood. 2013;122(18):3206–9.

14. Ceccaldi R, Parmar K, Mouly E, Delord M, Kim JM, Regairaz M, et al. Bone Marrow Failure in Fanconi Anemia Is Triggered by an Exacerbated p53/p21 DNA Damage Response that Impairs Hematopoietic Stem and Progenitor Cells. Cell Stem Cell. 2012;11(1):36–49.

15. Quentin S, Cuccuini W, Ceccaldi R, Nibourel O, Pondarre C, Pagès M-P, et al. Myelodysplasia and leukemia of Fanconi anemia are associated with a specific pattern of genomic abnormalities that includes cryptic RUNX1/AML1 lesions. Blood. 2011;117(15):161–70.

16. Wölfler A, Erkeland SJ, Bodner C, Valkhof M, Renner W, Leitner C, et al. A

functional single-nucleotide polymorphism of the G-CSF receptor gene predisposes individuals to high-risk myelodysplastic syndrome. Blood. 2005;105(9):3731–6.

17. West RR, Stafford DA, White AD, Bowen DT, Padua RA. Cytogenetic abnormalities in the myelodysplastic syndromes and occupational or environmental exposure. Blood. 2000;95(6):2093–7.

18. Mishra A, Anasetti C. Selection of Patients With Myelodysplastic Syndrome for Allogeneic Hematopoietic Stem Cell Transplantation. Clin Lymphoma Myeloma Leuk. 2016;16:49–52.

19. Cazzola M, Della Porta MG, Malcovati L. The genetic basis of myelodysplasia and its clinical relevance. Blood. 2013;122(25):4021–34.

20. Chen C-Y, Lin L-I, Tang J-L, Ko B-S, Tsay W, Chou W-C, et al. RUNX1 gene mutation in primary myelodysplastic syndrome - the mutation can be detected early at diagnosis or acquired during disease progression and is associated with poor outcome. Br J Haematol. 2007;139(3):405–14.

21. Delhommeau F, Dupont S, Valle V Della, James C, Trannoy S, Massé A, et al. Mutation in TET2 in Myeloid Cancers. N Engl J Med. 2009;360(22):2289–301.

22. Ernst T, Chase AJ, Score J, Hidalgo-Curtis CE, Bryant C, Jones A V, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat Genet. 2010;42(8):722–6.

23. Gelsi-Boyer V, Trouplin V, Roquain J, Adélaïde J, Carbuccia N, Esterni B, et al. ASXL1 mutation is associated with poor prognosis and acute transformation in chronic myelomonocytic leukaemia. Br J Haematol. 2010;151(4):365–75.

24. Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia. 2014;28(2):241–7.

25. Kosmider O, Gelsi-Boyer V, Cheok M, Grabar S, Della-Valle V, Picard F, et al. TET2 mutation is an independent favorable prognostic factor in myelodysplastic syndromes (MDSs). Blood. 2009;114(15):3285–91.

26. Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, et al. DNMT3A Mutations in Acute Myeloid Leukemia. N Engl J Med. 2010;363(25):2424–33.

27. Nikoloski G, Langemeijer SMC, Kuiper RP, Knops R, Massop M, Tönnissen ERLTM, et al. Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. Nat Genet. 2010;42(8):665–7.

28. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood. 2013;122(22):3616–27.

29. Steensma DP, Gibbons RJ, Mesa RA, Tefferi A, Higgs DR. Somatic point mutations in RUNX1/CBFA2/AML1 are common in high-risk myelodysplastic syndrome, but not in myelofibrosis with myeloid metaplasia. Eur J Haematol. 2005;74(1):47–53.

30. Greenberg P, Cox C, LeBeau MM, Fenaux P, Morel P, Sanz G, et al. International

Scoring System for Evaluating Prognosis in Myelodysplastic Syndromes. Blood. 1997;89(6):2079–88.

31. Kawabata H, Tohyama K, Matsuda A, Araseki K, Hata T, Suzuki T. Validation of the revised International Prognostic Scoring System in patients with myelodysplastic syndrome in Japan: results from a prospective multicenter registry. Int J Hematol. 2017;106(3):375–84.

32. Greenberg PL, Tuechler H, Schanz J, Sanz G, Garcia-Manero G, Solé F, et al. Revised international prognostic scoring system for myelodysplastic syndromes. Blood. 2012;120:2454–65.

33. Malcovati L, Germing U, Kuendgen A, Della Porta MG, Pascutto C, Invernizzi R, et al. Time-dependent prognostic scoring system for predicting survival and leukemic evolution in myelodysplastic syndromes. J Clin Oncol. 2007;25(23):3503–10.

34. Della Porta MG, Tuechler H, Malcovati L, Schanz J, Sanz G, Garcia-Manero G, et al. Validation of WHO classification-based Prognostic Scoring System (WPSS) for myelodysplastic syndromes and comparison with the revised International Prognostic Scoring System (IPSS-R). A study of the International Working Group for Prognosis in Myelodyspla. Leukemia. 2015;29(7):1502–13.

35. Smith C, Gasparetto M, Humphries K, Pollyea DA, Vasiliou V, Jordan CT. Aldehyde dehydrogenases in acute myeloid leukemia. Ann N Y Acad Sci. 2014;1310:58–68.

36. Gasparetto M, Smith CA. ALDHs in normal and malignant hematopoietic cells: Potential new avenues for treatment of AML and other blood cancers. Chem Biol Interact. 2017;276:46–51.

37. Venton G, Pérez-Alea M, Baier C, Fournet G, Quash G, Labiad Y, et al. Aldehyde dehydrogenases inhibition eradicates leukemia stem cells while sparing normal progenitors. Blood Cancer J. 2016;6(9):e469.

38. Deschler B, Lübbert M. Acute myeloid leukemia: Epidemiology and etiology. Cancer. 2006;107(9):2099–107.

39. Koppaka V, Thompson DC, Chen Y, Ellermann M, Nicolaou KC, Juvonen RO, et al. Aldehyde dehydrogenase inhibitors: a comprehensive review of the pharmacology, mechanism of action, substrate specificity, and clinical application. Pharmacol Rev. 2012;64(3):520–39.

40. M. A, Bagirova M, Nehir O, Yaman S, Sefik E, Cakir R, et al. Aldehyde Dehydrogenase: Cancer and Stem Cells. 2012.

41. Vasiliou V, Nebert DW. Analysis and update of the human aldehyde dehydrogenase (ALDH) gene family. Hum Genomics. 2005;2(2):138–43.

42. Marcato P, Dean CA, Giacomantonio CA, Lee PWK. Aldehyde dehydrogenase: Its role as a cancer stem cell marker comes down to the specific isoform. Cell Cycle. 2011;10(9):1378–84.

43. Ma I, Allan AL. The Role of Human Aldehyde Dehydrogenase in Normal and Cancer Stem Cells. Stem Cell Rev Reports. 2011;7(2):292–306.

44. Marchitti SA, Brocker C, Stagos D, Vasiliou V. Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. Expert Opin Drug Metab Toxicol. 2008;4(6):697–720.

45. Muzio G, Maggiora M, Paiuzzi E, Oraldi M, Canuto RA. Aldehyde dehydrogenases and cell proliferation. Free Radic Biol Med. 2012;52(4):735–46.

46. Moretti A, Li J, Donini S, Sobol RW, Rizzi M, Garavaglia S, et al. Crystal structure of human aldehyde dehydrogenase 1A3 complexed with NAD+ and retinoic acid. Sci Rep. 2016;6:35710.

47. Pors K, Moreb JS. Aldehyde dehydrogenases in cancer: an opportunity for biomarker and drug development? Drug Discov Today. 2014;19(12):1953–63.

48. Wu L, Lu X, Guo J, Zhang T, Wang F, Bao Y. Association between ALDH1L1 gene polymorphism and neural tube defects in the Chinese Han population. Neurol Sci. 2016;37(7):1049–54.

49. Jin S, Chen J, Chen L, Histen G, Lin Z, Gross S, et al. ALDH2(E487K) mutation increases protein turnover and promotes murine hepatocarcinogenesis. Proc Natl Acad Sci. 2015;112(29):9088–93.

50. Han H, Wang H, Yin Z, Jiang H, Fang M, Han J. Association of genetic polymorphisms in ADH and ALDH2 with risk of coronary artery disease and myocardial infarction: A meta-analysis. Gene. 2013;526(2):134–41.

51. Rizzo WB. Genetics and prospective therapeutic targets for Sjögren-Larsson Syndrome. Expert Opin Orphan Drugs. 2016;4(4):395–406.

52. Srivastava D, Singh RK, Moxley MA, Henzl MT, Becker DF, Tanner JJ. The Three-Dimensional Structural Basis of Type II Hyperprolinemia. J Mol Biol. 2012;420(3):176–89.

53. Pearl PL, Gibson KM, Acosta MT, Vezina LG, Theodore WH, Rogawski MA, et al. Clinical spectrum of succinic semialdehyde dehydrogenase deficiency. Neurology. 2003;60(9):1413–7.

54. Marcadier JL, Smith AM, Pohl D, Schwartzentruber J, Al-Dirbashi OY, Majewski J, et al. Mutations in ALDH6A1 encoding methylmalonate semialdehyde dehydrogenase are associated with dysmyelination and transient methylmalonic aciduria. Orphanet J Rare Dis. 2013;8(1):98.

55. Al Teneiji A, Bruun TUJ, Cordeiro D, Patel J, Inbar-Feigenberg M, Weiss S, et al. Phenotype, biochemical features, genotype and treatment outcome of pyridoxine-dependent epilepsy. Metab Brain Dis. 2016;32(2):443–51.

56. Sulem P, Gudbjartsson DF, Walters GB, Helgadottir HT, Helgason A, Gudjonsson SA, et al. Identification of low-frequency variants associated with gout and serum uric acid levels. Nat Genet. 2011;43(11):1127–30.

57. Baumgartner MR, Hu CA, Almashanu S, Steel G, Obie C, Aral B, et al. Hyperammonemia with reduced ornithine, citrulline, arginine and proline: a new inborn error caused by a mutation in the gene encoding delta(1)-pyrroline-5-carboxylate synthase. Hum Mol Genet. 2000;9(19):2853–8.

58. Tomita H. Aldehyde dehydrogenase 1A1 in stem cells and cancer. Oncotarget. 2016;7(10):11018–32.

59. Kashii-Magaribuchi K, Takeuchi R, Haisa Y, Sakamoto A, Itoh A, Izawa Y, et al. Induced Expression of Cancer Stem Cell Markers ALDH1A3 and Sox-2 in Hierarchical Reconstitution of Apoptosis-resistant Human Breast Cancer Cells. Acta Histochem Cytochem. 2016;49(5):149–58.

60. Flahaut M, Jauquier N, Chevalier N, Nardou K, Balmas Bourloud K, Joseph J-M, et al. Aldehyde dehydrogenase activity plays a Key role in the aggressive phenotype of neuroblastoma. BMC Cancer. 2016;16(1):781.

61. Müller MF, Zhou Y, Adams DJ, Arends MJ. Effects of long-term ethanol consumption and Aldh1b1 depletion on intestinal tumourigenesis in mice. J Pathol. 2016;241(5):649–60.

62. Mudannayake JM, Mouravlev A, Fong DM, Young D. Transcriptional activity of novel ALDH1L1 promoters in the rat brain following AAV vector-mediated gene transfer. Mol Ther Methods Clin Dev. 2016;3:16075.

63. Strickland KC, Krupenko NI, Dubard ME, Hu CJ, Tsybovsky Y, Krupenko SA. Enzymatic properties of ALDH1L2, a mitochondrial 10-formyltetrahydrofolate dehydrogenase. Chem Biol Interact. 2011;191(1–3):129–36.

64. You Q, Guo H, Xu D. Distinct prognostic values and potential drug targets of ALDH1 isoenzymes in non-small-cell lung cancer. Drug Des Devel Ther. 2015;9:5087–97.

65. Van Wassenhove LD, Mochly-Rosen D, Weinberg KI. Aldehyde dehydrogenase 2 in aplastic anemia, Fanconi anemia and hematopoietic stem cells. Mol Genet Metab. 2016;119(1–2):28–36.

66. Budas GR, Disatnik M-H, Chen C-H, Mochly-Rosen D. Activation of aldehyde dehydrogenase 2 (ALDH2) confers cardioprotection in protein kinase C epsilon (PKCε) knockout mice. J Mol Cell Cardiol. 2010;48(4):757–64.

67. Buchman CD, Mahalingan KK, Hurley TD. Discovery of a series of aromatic lactones as ALDH1/2-directed inhibitors. Chem Biol Interact. 2015;234:38–44.

68. Holmes RS. Biochemical Genetics of Opossum Aldehyde Dehydrogenase 3: Evidence for Three ALDH3A-Like Genes and an ALDH3B-Like Gene. Biochem Genet. 2010;48(3–4):287–303.

69. Gross A, Ong TR, Grant R, Hoffmann T, Gregory DD, Sreerama L. Human aldehyde dehydrogenase-catalyzed oxidation of ethylene glycol ether aldehydes. Chem Biol Interact. 2009;178(1–3):56–63.

70. Kitamura T, Naganuma T, Abe K, Nakahara K, Ohno Y, Kihara A. Substrate specificity, plasma membrane localization, and lipid modification of the aldehyde dehydrogenase ALDH3B1. Biochim Biophys Acta - Mol Cell Biol Lipids. 2013;1831(8):1395–401.

71. Yagishita A, Ueno T, Esumi H, Saya H, Kaneko K, Tsuchihara K, et al. Development of Highly Selective Fluorescent Probe Enabling Flow-Cytometric Isolation of ALDH3A1-Positive Viable Cells. Bioconjug Chem. 2016;28(2):302–306.

72. Chen J, Shin VY, Siu MT, Ho JCW, Cheuk I, Kwong A. miR-199a-5p confers tumor-suppressive role in triple-negative breast cancer. BMC Cancer. 2016;16(1):887.

73. Pemberton TA, Srivastava D, Sanyal N, Henzl MT, Becker DF, Tanner JJ. Structural Studies of Yeast Δ 1 -Pyrroline-5-carboxylate Dehydrogenase (ALDH4A1): Active Site Flexibility and Oligomeric State. Biochemistry. 2014;53(8):1350–9.

74. Kardos GR, Wastyk HC, Robertson GP. Disruption of Proline Synthesis in Melanoma Inhibits Protein Production Mediated by the GCN2 Pathway. Mol Cancer Res. 2015;13(10):1408–20.

75. Hilvo M, de Santiago I, Gopalacharyulu P, Schmitt WD, Budczies J, Kuhberg M, et al. Accumulated Metabolites of Hydroxybutyric Acid Serve as Diagnostic and Prognostic Biomarkers of Ovarian High-Grade Serous Carcinomas. Cancer Res. 2016;76(4):796–804.

76. Luo M, Tanner JJ. Structural Basis of Substrate Recognition by Aldehyde Dehydrogenase 7A1. Biochemistry. 2015;54(35):5513–22.

77. Ding K, Shameer K, Jouni H, Masys DR, Jarvik GP, Kho AN, et al. Genetic Loci Implicated in Erythroid Differentiation and Cell Cycle Regulation Are Associated With Red Blood Cell Traits. Mayo Clin Proc. 2012;87(5):461–74.

78. Grinberg M, Stöber RM, Edlund K, Rempel E, Godoy P, Reif R, et al. Toxicogenomics directory of chemically exposed human hepatocytes. Arch Toxicol. 2014;88(12):2261–87.

79. Henrion MYR, Purdue MP, Scelo G, Broderick P, Frampton M, Ritchie A, et al. Common Variation at 1q24.1 (ALDH9A1) Is a Potential Risk Factor for Renal Cancer. PLoS One. 2015;10(3):e0122589.

80. Hanna MC, Blackstone C. Interaction of the SPG21 protein ACP33/maspardin with the aldehyde dehydrogenase ALDH16A1. Neurogenetics. 2009;10(3):217–28.

81. Holmes RS. Comparative and evolutionary studies of ALDH18A1 genes and proteins. Chem Biol Interact. 2016;276:2–8.

82. Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, Bouvard V, et al. Carcinogenicity of alcoholic beverages. Lancet Oncol. 2007;8(4):292–3.

83. Chute JP, Muramoto GG, Whitesides J, Colvin M, Safi R, Chao NJ, et al. Inhibition of aldehyde dehydrogenase and retinoid signaling induces the expansion of human hematopoietic stem cells. Proc Natl Acad Sci U S A. 2006;103(31):11707–12.

84. Muramoto GG, Russell JL, Safi R, Salter AB, Himburg HA, Daher P, et al. Inhibition of Aldehyde Dehydrogenase Expands Hematopoietic Stem Cells with Radioprotective Capacity. Stem Cells. 2010;28(3):523–34.

85. Gasparetto M, Sekulovic S, Zakaryan A, Imren S, Kent DG, Humphries RK, et al. Varying levels of aldehyde dehydrogenase activity in adult murine marrow hematopoietic stem cells are associated with engraftment and cell cycle status. Exp Hematol. 2012;40(10):857–66.

86. Collins SJ. Retinoic acid receptors, hematopoiesis and leukemogenesis. Curr Opin

Hematol. 2008;15(4):346–51.

87. van den Hoogen C, van der Horst G, Cheung H, Buijs JT, Lippitt JM, Guzman-Ramirez N, et al. High Aldehyde Dehydrogenase Activity Identifies Tumor-Initiating and Metastasis-Initiating Cells in Human Prostate Cancer. Cancer Res. 2010;70(12):5163–73.

88. Vasiliou V, Thompson DC, Smith C, Fujita M, Chen Y. Aldehyde dehydrogenases: From eye crystallins to metabolic disease and cancer stem cells. Chem Biol Interact. 2013;202(1–3):2–10.

89. Luo Y, Dallaglio K, Chen Y, Robinson WA, Robinson SE, McCarter MD, et al. ALDH1A isozymes are markers of human melanoma stem cells and potential therapeutic targets. Stem Cells. 2012;30(10):2100–13.

90. Patel M, Lu L, Zander DS, Sreerama L, Coco D, Moreb JS. ALDH1A1 and ALDH3A1 expression in lung cancers: Correlation with histologic type and potential precursors. Lung Cancer. 2008;59(3):340–9.

91. Marcato P, Dean CA, Pan D, Araslanova R, Gillis M, Joshi M, et al. Aldehyde Dehydrogenase Activity of Breast Cancer Stem Cells Is Primarily Due To Isoform ALDH1A3 and Its Expression Is Predictive of Metastasis. Stem Cells. 2011;29(1):32–45.

92. Chen Y, Orlicky DJ, Matsumoto A, Singh S, Thompson DC, Vasiliou V. Aldehyde dehydrogenase 1B1 (ALDH1B1) is a potential biomarker for human colon cancer. Biochem Biophys Res Commun. 2011;405(2):173–9.

93. Singh S, Arcaroli JJ, Orlicky DJ, Chen Y, Messersmith WA, Bagby S, et al. Aldehyde Dehydrogenase 1B1 as a Modulator of Pancreatic Adenocarcinoma. Pancreas. 2016;45(1):117–22.

94. Marchitti SA, Orlicky DJ, Brocker C, Vasiliou V. Aldehyde dehydrogenase 3B1 (ALDH3B1): immunohistochemical tissue distribution and cellular-specific localization in normal and cancerous human tissues. J Histochem Cytochem. 2010;58(9):765–83.

95. Ginestier C, Wicinski J, Cervera N, Monville F, Finetti P, Bertucci F, et al. Retinoid signaling regulates breast cancer stem cell differentiation. Cell Cycle. 2009;8(20):3297–302.

96. Rice KL, Izon DJ, Ford J, Boodhoo A, Kees UR, Greene WK. Overexpression of stem cell associated ALDH1A1, a target of the leukemogenic transcription factor TLX1/HOX11, inhibits lymphopoiesis and promotes myelopoiesis in murine hematopoietic progenitors. Leuk Res. 2008;32(6):873–83.

97. Wang H, Zhou H, Mahler S, Chervenak R, Wolcott M. Alcohol Affects the Late Differentiation of Progenitor B Cells. Alcohol Alcohol. 2011;46(1):26–32.

98. Latvala J, Parkkila S, Niemelä O. Excess alcohol consumption is common in patients with cytopenia: studies in blood and bone marrow cells. Alcohol Clin Exp Res. 2004;28(4):619–24.

99. Yin S. The role of aldehyde dehydrogenase (ALDH) isoform 1A3 in the pathogenesis

of human acute myeloid leukemia (AML). 2011.

100. Voulgaridou G, Kiziridou M, Mantso T, Chlichlia K, Galanis A, Koukourakis M, et al. Aldehyde dehydrogenase 3A1 promotes multi-modality resistance and alters gene expression profile in human breast adenocarcinoma MCF-7 cells. Int J Biochem Cell Biol. 2016;77(PtA):120–8.

101. Basak N, Banerjee S. Mitochondrial dependency in progression of acute myeloid leukemia. Mitochondrion. 2015;21:41–8.

102. Gupta M, Madkaikar M, Rao V, Mishra A, Govindaraj P, Thangaraj K, et al. Mitochondrial DNA variations in myelodysplastic syndrome. Ann Hematol. 2013;92(7):871–6.

103. Schildgen V, Wulfert M, Gattermann N. Impaired mitochondrial gene transcription in myelodysplastic syndromes and acute myeloid leukemia with myelodysplasia-related changes. Exp Hematol. 2011;39(6):666–75.