ASSESSING THE ROLE OF THE TRANSCRIPTION FACTOR CTCF IN THE PATHOGENESIS OF MYELODYSLASTIC SYNDROMES

Background and Rationale

Myelodysplastic Syndromes (MDS) are clonal hematopoietic disorders that arise de novo (primary MDS) or after radiation/chemotherapy (secondary or therapy-related MDS). MDS incidence is unclear due to differences in diagnostic sensitivity and variation in data collection as well as in the mean-age reference population. Environmental risk factors and/or genetic predispositions possibly influence MDS incidence in certain geographical areas. To date, the overall incidence rate is estimated to be 10 new cases per 100,000 people per annum. Particularly, primary MDS show increased incidence with ageing, frequently affecting people over 70. Due to increased life expectancy and exposure to environmental risk factors, such as industrial fumes and pesticides, together with extensive use of chemo- or radiotherapy for the treatment of cancers, MDS incidence is likely to increase in the next years, becoming a global health problem.

MDS pathogenesis has not been elucidated so far and, unlike for other haematological malignancies, no clear molecular and cytogenetic alterations have been associated to MDS subtypes (apart from the 5q-syndrome) or to different stages of disease. MDS represent a heterogeneous group of diseases characterized by clonal proliferation, bone marrow failure and increased risk of transformation into Acute Myeloid Leukemia (AML) (1).

Recent data describe the occurrence of mutations in genes involved in the epigenetic regulation of gene transcription. Among recurrent alterations, IDH1/2, TET2, DNMT3A, and EZH2 mutations appear to affect DNA and/or histone lysine methylation (2). Moreover, promoter DNA hypermethylation has been described as a mechanism of transcriptional silencing in MDS (3), and hypermethylation phenotype of specific genes has been associated with MDS rapid progression to AML (4). Such epigenetic complexity, described in MDS by genome-wide analyses of DNA methylation (5, 6), provide the rationale for the usage of hypomethylating agents in the treatment of MDS (7). Indeed, treatment options for MDS management range from supportive care to aggressive treatment, which may slowdown or prevent disease progression. The hypomethylating agents azacitidine and decitabine achieve significant overall response rates in MDS, and improve outcome of higher risk patients. However, the results obtained with hypomethylating agents are not considered satisfactory, and efforts are being made to improve these results, by testing novel therapy schedules or administration routes. Also, research efforts are ongoing to better understand the mechanism of hypomethylating agents’ action, to optimize their effect, to improve drug exposure while reducing side effects, and to discover synergic combination therapies.
DNA methylation, other than silencing promoters, could also prevent the binding of transcription factors to other regulatory DNA sequences. CTCF (CCCTC-binding factor) is a highly conserved multifunctional protein with 11 zinc fingers, able to recognize different DNA sites (8). CTCF carries out many activities in transcriptional regulatory processes, such as promoter activation/repression, enhancer blocking and/or barrier insulation, genomic imprinting, and long-range chromatin interactions. Moreover, a recent study reported an intragenic role for CTCF that associates DNA methylation with alternative splicing. Exon inclusion into mature mRNA could be favoured by CTCF binding to intragenic target sites responsible for the elongation rates reduction. DNA methylation of CTCF intragenic sites avoids CTCF binding and favours exon exclusion from mature mRNA (9). CTCF is a nuclear protein with a ubiquitous expression and about 30,000 DNA target sites (8). Although essential for cell survival, CTCF expression levels and nuclear distribution patterns are tissue-specific and differ during cell cycle phases. Of note, variations in CTCF expression levels may influence proliferation and differentiation in vitro (10). It has been reported that during the differentiation of the multipotent cell line K562 to erythroid lineage, CTCF down-regulation significantly inhibits cell differentiation (11).

CTCF was found to be located at human chromosome 16q22 within the loss of heterozygosity region associated with breast and prostate cancer (12, 13). However, mutations leading to a complete loss of CTCF function in the retained allele were not identified (14). In solid neoplasms, the observed CTCF mutations have been shown to selectively hamper its DNA binding specificity without completely abrogating its function (14). Therefore, it is conceivable that epigenetic mechanisms could be able to disrupt the spectrum of CTCF target specificities contributing to the development of a malignant phenotype. To date, the relationship between CTCF functional alteration and the onset of haematological malignancies has never been investigated.

Aim of the study
The present research project aims at improving our understanding of the pathogenetic mechanisms underlying MDS onset, by integrating genomic, transcriptomic and epigenetic approaches. The molecular events regulating the clinical heterogeneity of MDS have not been elucidated so far, and the application and integration of recently introduced molecular technologies will help us to understand MDS clinic, thus improving our treatment options and refining prognostic models for MDS patients. Particularly, we will focus on a transcription factor of broad activity, namely CTCF (CCCTC-binding factor), whose dysregulation has been often described in non-hematologic malignancies. We will explore the role of CTCF in the onset of MDS by evaluating CTCF mutations and its possible involvement in chromosomal rearrangements. Furthermore, taking
advantage of chromatin immunoprecipitation coupled with deep-sequencing technologies (ChIP-Seq), we will investigate variations in the global distribution of CTCF DNA-binding sites and analyze changes in genome-wide DNA methylation (MeDIP-Seq). Finally, the integration of ChIP-Seq, MeDIP-Seq and Gene expression profiling (GEP) data will allow us to identify gene pathways regulated by CTCF and potentially involved in the MDS pathogenesis.

**Methods**

Firstly, the role of CTCF in MDS onset will be explored by studying its involvement in chromosomal rearrangements, and assessing CTCF expression at RNA and protein levels. Furthermore, variations in the global distribution of CTCF DNA-binding sites will be investigated and compared with changes in genome-wide DNA methylation. Finally, CTCF target genes with altered expression will be identified suggesting new molecular pathways that are dysregulated in MDS.

The occurrence of CTCF gene mutations will be investigated by massively parallel. This approach is able to improve the sensitivity of genetic variants detection (1-2% frequency) as compared with Sanger-based sequencing (20% frequency). CTCF mutations will be analyzed from bone marrow (BM) DNA of about 100 patients with MDS at diagnosis.

A Fluorescent In Situ Hybridization (FISH) assay with specific home-brew probes will be developed to identify chromosomal rearrangements of the CTCF locus in MDS patients. The home-brew probes will consist of Bacterial Artificial Chromosomes (BAC) selected according to the University of California Santa Cruz database (UCSC; http://genome.ucsc.edu; February 2009 release) that can entirely encompass the genomic CTCF sequence. The mapping position of selected BAC probes will be verified on normal control metaphases; at least 200 cells will be examined in each experiment. Digital images will be obtained using a Leica DMRXA epifluorescence microscope equipped with a cooled CCD camera. Specific signals will be recorded separately as gray-scale images, while image pseudocoloring and merging will be performed with Adobe Photoshop software. Once validated, CTCF FISH will be performed on BM cells from MDS patients at diagnosis, as a tool to unveil possible CTCF deletions, amplifications, and involvements in chromosomal translocations, inversions or insertions. About 100 MDS patients, diagnosed according to the 2008 World Health Organization (WHO) criteria, will be analyzed.

CTCF expression will be assessed by quantitative Real Time PCR (qRT-PCR) in different subgroups of MDS patients, namely refractory anemia (RA), refractory anemia with ringed sideroblasts (RARS), refractory anemia with excess blasts (RAEB-1, RAEB-2) and MDS associated with isolated del(5q). qRT-PCR experiments will be performed using the 1X Platinum®SYBR Green qPCR SuperMix-UDG run on the ABI Prism 7300 Sequence Detection System. β-
glucuronidase (β-GUS) gene will be used as endogenous control and all qRT-PCR data will be compared to a pool of cDNA samples derived from 10 healthy controls (calibrator). The DataAssist software tool (Applied Biosystems) will be employed to analyze the relative expression results. Moreover, the expression level of two known CTCF isoforms will be examined by qRT-PCR with specific primer pairs. Finally, MDS patients with CTCF gene dysregulation will be analyzed by Western Blot to quantify CTCF expression at protein level.

Subsequently, genome-wide analyses will be performed in order to establish CTCF DNA-binding site maps in healthy subjects and MDS patients by ChIP-seq, a technique that couples chromatin immunoprecipitation with next generation sequencing (ChIP-seq). These experiments will be carried out on primary magnetically-isolated CD34+ cells from MDS BM-derived mononuclear cells (BM-MNC). ChIP analyses will be performed using the EZ-ChIP kit (Millipore). Briefly, cells will be firstly fixed with formaldehyde to covalently crosslink proteins to DNA. After cross-linking, the chromatin will be fragmented into pieces of about 150 to 500 bp by sonication. The average fragment size will be empirically confirmed by electrophoresis, and DNA will be immunoprecipitated using a CTCF specific antibody followed by Protein G-conjugated agarose beads as the secondary reagent. Then, reverse crosslinks of protein-DNA complexes will be carried out by a 65°C incubation. DNA will be subjected to deep sequencing and ChIP-Seq data will be analyzed by the open-source Galaxy tool (http://main.g2.bx.psu.edu/). A typical ChIP-Seq experiment generates millions of short (35-75 bp) DNA sequences, representing the ends of the immunoprecipitated DNA fragments. All sequences will be aligned to the reference genome (hg19), and a peak-calling algorithm, able to produce a list of regions with high read densities (peaks), will be applied. An input DNA control will be included to improve peak calling accuracy and eliminate artifacts. ChIP-Seq experiments for the identification of CTCF DNA-binding sites will be also performed in MDS patients treated with hypomethylating agents, in order to evaluate therapy-induced CTCF DNA-binding site changes and their possible correlations with the clinical response. These experiments would shed light on the mechanisms by which hypomethylating drugs act while providing information about the MDS pathogenesis at epigenomic level. Bioinformatic analysis of CTCF road maps produced by ChIP-Seq experiments will be carried out to define the relationship between the functional role of CTCF and the MDS pathogenesis.

To generate high-resolution maps of the methylome, DNA of MDS patients will be fragmented and regions of methylated DNA will be immunoprecipitated using 5-methylcytosine-specific antibodies (MeDIP-Seq). Sequencing of these fragments will identify genomic methylation-enriched regions. The obtained loci will be compared with genome-wide maps for CTCF binding sites in different subtypes of MDS patients, in order to verify whether the different distribution of the CTCF binding
sites is due to their hypermethylation. Finally, MDS patients showing a peculiar distribution of CTCF binding will be studied by Gene expression profiling (GEP). This analysis will be useful to correlate probable alteration of genome-wide CTCF binding sites to specific gene signatures and will show if the global state of chromatin structure is able to modify gene expression. Briefly, RNA will be isolated using the QIAcube (Qiagen) system and spectrophotometrically quantified. RNA purity and quality will be evaluated using the Agilent Bioanalyzer 2100 and an Agilent exon array will be employed to obtain a global picture of RNA expression. To date, MDS have not been analyzed extensively by GEP due to heterogeneity of cell populations, nonetheless distinct GEP were associated with specific FAB and cytogenetic subclasses (RA/RARS; 5q-, trisomy 8 or -7/del(7q)), and with different stages of disease, particularly early and advanced MDS. Overall, GEP analyses on a large cohort of MDS samples revealed that the most deregulated gene pathways were related to interferon and thrombopoietin signaling and to the Wnt pathway.

**Expected results**

The integration of ChiP-Seq, MeDIP-Seq and GEP, allowing us to identify gene pathways regulated by CTCF, may have important impacts on the comprehension of MDS pathogenesis. Moreover, the interaction between genes and environment may be important in influencing the epigenome, which is less stable than the genome, and varies with age and disease. In this respect, DNA methyloma study of MDS patients could shed light on the role played by ageing and environmental factors on the onset of these haematological malignancies, as well as clarify the mechanisms by which hypomethylating drugs act. Indeed, we expect to unveil novel target genes to be exploited as new molecular markers for diagnosis, prognosis, as well as for targeted therapy design.
Bibliography