This final report outlines the progress we have made in achieving the goals set out in the Young Investigator Award proposal entitled "Genomic Instability in MDS". Significant progress has been made in the objectives of both Specific Aims; however, the greatest advancement has occurred in the characterization of chromosomal abnormalities in MDS using high-density SNP arrays. Therefore, we focus on those results in this report.

Table 1. General characteristics of MDS patients (n=129)

Characteristic	Value (%)
Age -yr	
Median	67
Range	19 - 86
Sex	
Male	80 (62)
Female	49 (38)
IPSS risk category	
Low	29 (22)
Int-1	41 (32)
Int-2 or high	51 (40)
Unclassified	8 (6)
WHO classification	
low grade MDS	
RA/RCMD	15 (12)
RARS/RCMD-RS	14 (11)
5q- syndrome	2 (2)
MDS-U	14 (11)
Advanced MDS	
RAEB1/2	25 (19)
AML	24 (19)
CMML-2	10 (8)
MDS/MPD overlap	
CMML-1	11 (9)

SNP array-based karyotyping improves cytogenetic diagnosis and clinical prognosis of myelodysplastic syndrome. We have applied 250K high-density SNP arrays (SNP-A) to study chromosomal lesions in bone marrow samples from 129 MDS patients (Table 1). Using SNP-A, chromosomal aberrations were found in 75% of MDS patients (vs. 52% by metaphase chromosome (MC) exam, which was not informative in an additional 6% of cases (Figure 1A). Moreover, previously unrecognized lesions were detected in both patients with normal MC, as well as in those with known lesions. More chromosomal aberrations were revealed, with a greater proportion of patients showing >1 defect (27/129 vs. 64/129, $p<0.001$, Figure 1B). Most significantly, segmental uniparental disomy (UPD) was detected in 25% of patients, a lesion resulting in gene copy-neutral loss of heterozygosity that cannot be detected by MC. When patient groups with normal MC or SNP-A results were compared, a lower proportion of patients with low-grade disease was found if the karyotype was assessed to be normal using the new technology, consistent with an overall negative clinical impact of newly identified lesions. Separate analysis of patients with normal and non-informative MC showed that SNP-A can identify chromosomal aberrations in 65% and 50% of these patients, respectively (Figure 2).

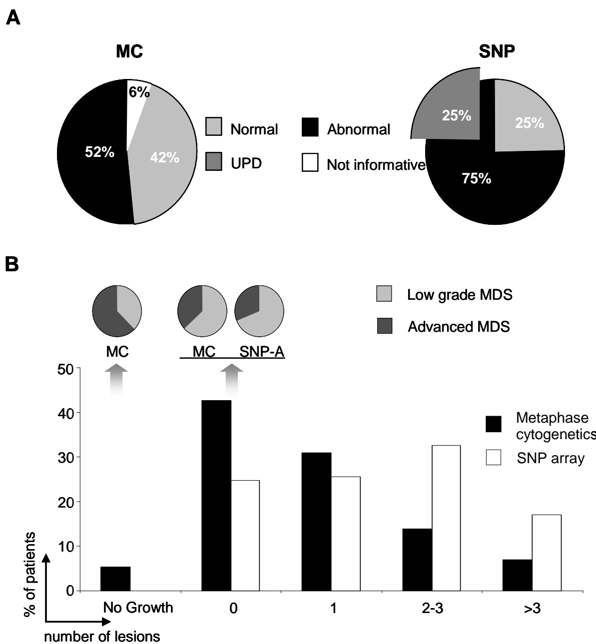

Figure 1. Sensitivity of SNP karyotyping and MC for detection of chromosomal abnormalities.

The clinical relevance of new lesions remains to be established. As patients with normal MC show significant clinical diversity, we analyzed the survival by dividing this heterogenous group into two categories: one with normal karyotype according to both MC and SNP-A and the other comprised of patients with normal MC in whom new lesions were identified using SNP-A. The latter showed impaired survival compared to those with normal MC and SNP-A karyotype

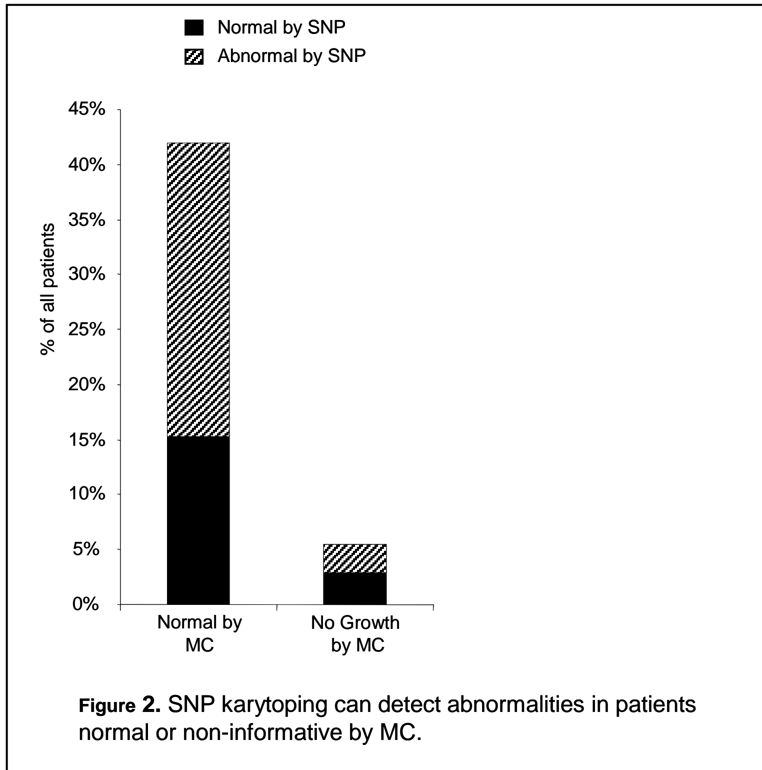

Figure 2. SNP karytoping can detect abnormalities in patients normal or non-informative by MC.

(median survival 8.4 vs. 21.5 months, p=0.003).We have also tested whether the newly identified lesions have a potential to modify the currently accepted IPSS system. We have grouped patients based on their risk category including IPSS-low, Int-1, Int-2, and high-risk groups. For each risk group patients were subdivided into those in whom SNP-A did not detect additional aberrations and those in whom new lesions were found. The survival curves clearly separated for patients originally classified as IPSS Int-1 and Int-2, suggesting that SNP-A conveys additional information allowing for better prognostic resolution. Due to a short observation period and variable impact of individual lesions, the survival within the low-risk group did not differ between SNP-A-confirmed karyotypes and those in whom MC was modified.

The findings generated in this study are under consideration at PLOS1.

Identification of chromosomal abnormalities in healthy bone marrow using 250K SNP arrays. Myelodysplastic syndromes (MDS) are characterized by the presence of clonal chromosomal abnormalities detectable by traditional cytogenetics in around 50% of patients. We have demonstrated that a higher percentage of unbalanced clonal chromosomal defects and uniparental disomy (UPD) can be identified in MDS using high-density SNP arrays. The higher detection rate may have important clinical consequences. However, such findings must be considered in the context of normal karyotypic variation. Before the clinical relevance of new lesions identified by SNP-A can be presumed, several issues must be addressed. i) The increased precision of karyotypic analysis may lead to the detection of lesions in normal bone marrow. This is especially relevant to elderly patients with MDS for whom adequate age-matched comparisons should be performed. ii) The normal distribution of chromosomal changes across the genome must be defined as it may overlap with that in disease. iii) The minimal clonal size detectable by SNP-A analysis is of importance, as hematopoiesis may be oligo- rather than monoclonal in many conditions. We stipulated that the clinical applicability of SNP-A-based karyotyping will depend upon the findings in healthy controls and have studied 51 normal bone marrows using SNP-A karyotype analysis. The age of the normal controls was slightly lower than that of our MDS patient cohort (average 38 years, range 15-76 years). Lesions were identified in 78% (40/51) of the controls.

The high level of chromosomal lesions in putatively healthy individuals is not entirely unexpected, as it has recently become clear that the human genome harbors a number of regions of copy number polymorphism (CNP). In our cohort we identified 14 regions of CNP, all of which have been previously reported. When these regions were excluded, the percent of healthy individuals with chromosomal abnormalities dropped to 65%. 14 marrows contained 1

chromosomal change, while 2-4 were found in 19 marrows. Both loss and gain of sequences were detected. The size of the largest deletion and duplication was 2.35 Mb and 1.64 Mb, respectively. Karyotypic abnormalities identified in control samples appeared to be randomly distributed across the genome. No lesions were identified on chromosome 8 or chromosome 5q, although one deletion on chromosome 5p was found. A small deletion of 7q was detected (1.2 Mb); overall, the regions frequently affected in MDS were rarely altered in controls. Finally, we identified loss of heterozygosity (LOH) due to UPD in 4 control samples. One sample had LOH at 4 loci (11q12.3-11q13.5, 20p12.3-20p12.1, 3q13.12-3q13.3, 9q21.33-9131.3) while another harbored two regions of LOH (4q34.1-4q25.1, 9p24.1-9p23) and two contained one area of LOH (13q21.3-13q32.1; 13q14.11-13q21.31).

To assess the minimal detectable clonal size, we next examined the sensitivity of SNP-A analysis to the admixture of normal cells. When samples(deletion of 7q, trisomy 8) or cell lines (trisomy 21) identified as abnormal by traditional cytogenetics were serially diluted with normal genomic DNA, the previously identified lesions could still be detected when the sample contained 25% normal genomic DNA, but at 50% the copy number analysis appeared normal. Our result underscores that, only significantly expanded clones can be systematically detected by SNP-A analysis. Pathogeneic defects have to exclude regions of CNP and must be larger than changes in healthy controls. Additionally, they would preferentially occur in regions not frequently affected in normal bone marrow. Our studies reveal the physiologic level of chromosomal abnormalities present in healthy controls and allow us to design criteria for defining abnormal karyotypes as measured by SNP-A.

The studies outlined here were presented at the annual American Society of Hematology meeting (O'Keefe CL, Gondek LP, Tiu R, Nearman ZP, Wlodarski M, Nannya Y, Ogawa S and Maciejewski JP. Identification of chromosomal abnormalities in healthy bone marrow using 250K SNP arrays. Blood 2006; 108: 588a).

Copy number variants as a part of the complex genetic traits predisposing to bone marrow failure. Individual variability, including disease susceptibility, is determined by the interaction of inherited single base differences (single nucleotide polymorphisms, SNPs) and copy number variants (CNVs) of large genomic regions. A complex combination of these factors may result in a genetic background predisposing to disease. Regions of CNV account for approximately 12% of the human genome, including coding sequences and can range in size from kilobases to megabases. Recent studies have investigated the correlation between CNVs and complex conditions, including mental retardation, lupus and cardiovascular disease. While SNPs have been intensely investigated in many diseases, the influence of CNVs on disease susceptibility is only poorly understood. With the advent of high-throughput, high density array technology, global analysis of complex disease predisposition traits, including CNVs, can be performed.

We have applied high-density SNP arrays (SNP-A) for the analysis of somatic chromosomal defects in various hematologic disorders. During our studies we noted a high frequency of germ-line CNVs, complicating our analysis of somatic defects. This observation lead us to the hypothesis that CNVs can themselves constitute predisposition factors to disease and chose to systematically investigate their type and frequency in myeloid disorders including aplastic anemia (AA; N=65), myelodysplastic syndrome (MDS; N=145) and primary and secondary (non-core binding factor) acute myeloid leukemia (AML; N=75). We performed whole genome scanning in patients and a cohort of healthy controls (N=79) using the Affymetrix 250K SNP array.

We first identified and catalogued CNVs in controls; their frequency was compared to those reported in the Database of Genomic Variants (http://projects.tcag.ca/variation/) and found to be similar. The CNVs ranged in size from 245.6 Kb to 2.32 Mb (average 805.9 Kb) and were identified on all chromosomes except 5, 13, 16, 18 and 21. We next analyzed copy

number changes in patients with myeloid disorders. Using controls (both our cohort and those in the literature) as a reference we determined the frequencies of recurrent CNVs in patients. For most of the CNVs the frequency was <10% within the individual patient groups, similar to what was seen in controls. Nonetheless, four regions (2 distinct loci in the pericentromeric region of 14q, pericentromeric 15q and a locus on 17q21.31) were identified in over 15% of samples studied. We then determined whether a distinct CNV is associated with specific disease risk. While for most CNVs the frequencies found in patients were similar to those in controls, two regions, 3q29 and 14q11.2, were more frequently encountered in patients with AML (3q29, 27/75 vs. 13/79 in controls, p=0.01; 14q11.2, 20/75 vs. 8/79 in controls, p=0.014). The region at 3q29 contains several genes and is a common breakpoint region for hematologic malignancies including MDS and AML, suggesting that this chromosomal area sensitive to physical rearrangement. The locus at 14q11.2 is a known hypervariable region, containing T cell receptor genes. In sum, in addition to SNPs, CNVs may be a part of complex genetic traits in patients with AA, MDS and AML and constitute disease predisposition factors. Beyond their potential role in disease, CNVs have to be excluded in SNP array-based analysis of somatic chromosomal lesions.

The studies outlined above were presented as an oral presentation at the annual American Society of Hematology meeting (O'Keefe CL, Gondek L, Wlodarski M, Karp J, McDevitt M and Maciejewski JP. Can genomic copy number variants be a part of complex genetic traits predisposing to marrow failure? Blood 2007 110: 39a).

Newly discovered lesions, including uniparental disomy, have prognostic impact. We have analyzed an additional 174 patients with MDS, MDS/MPD, or MDS-derived AML using 250K SNP arrays. Smaller previously unknown cryptic chromosomal aberrations were identified, as well as larger defects found by metaphase cytogenetics (MC). These defects appeared in bone marrow cells (as opposed to nonclonal lymphocytes) and in some instances were limited to one gene; they therefore likely represented defects with pathologic significance. The results of confirmatory TaqMan PCR studiesand microsatellite analysis (N=7) provided independent validation of the SNP-A findings. Confirmation of the somatic nature of SNP-A–detected lesions was performed using paired SNP-A analysis applied to both bone marrow and sorted nonclonal CD3+ cells (n =5); confirmation included UPD (6)(p21.2-pter), UPD (11)(q13.5-qter), UPD (4)(q23- qter), UPD (7)(q11.23-qter) and del (7)(q22.1). Such segmental UPD was most commonly found in chromosomal regions frequently identified as abnormal in MDS by MC and was not present in nonclonal lymphocytes, suggesting a somatic origin of SNP-A–detected lesions. UPD, deletions and duplications involved several important chromosomal regions affected in multiple patients

SNP-A facilitates precise mapping of lesions; LOH affecting common regions due to deletion and UPD can be grouped to better define boundaries of minimal overlapping regions. For demonstration purposes, we focused here on specific lesions on chromosome 7 and their prognostic significance. When we compared the survival of 3 groups of patients—one with normal SNP-A karyotype, another with previously known deletion 7/7q, and another with normal MC and new cryptic lesions (including deletions and UPD) on chromosome, those with previously known and new abnormalities showed a comparably poor prognosis, distinct from those with a normal karyotype by both MC and SNP-A (median survival 6 vs 8 vs 39 months, respectively, p<.002). When the group of patients with UPD7q (3 patients with normal MC, one with +13 and one with -Y) was analyzed separately, shorter survival was observed (median survival 3 vs 39 months, p<.001), indicating the clinical relevance of SNP-A and UPD7.

Overall, SNP-A allowed for identification of chromosomal defects in 78% of MDS patients, as compared with 59% by MC. Similarly, a higher detection rate of lesions was found in patients with MDS/MPD and sAML at 75% versus 37%, and 77% versus 53% for SNP-A and MC, respectively. SNP-A does not rely on cell growth, and SNP-A testing was successful

in all patients. UPD as the sole or a concurrent defect was detected in 20%, 35%, and 23% of patients with MDS, MDS/MPD, and sAML, respectively. More chromosomal aberrations were revealed by SNP-A as compared with MC, with a greater proportion of patients showing more than 1 defect (p<.001). As a result, when patient groups with normal MC or SNP-A results were compared, fewer patients with low-grade disease were found if the karyotype was assessed to be normal using the new technology, consistent with an overall negative clinical impact of newly identified lesions. In general, MDS patients in whom multiple lesions were detected were more likely to have a more advanced disease and show a more rapid increase in blasts. For example, in a patient with MDS/MPD-U, evolution to sAML was associated with occurrence of additional lesions such as UPD6p. Separate analysis of patients with normal or non-informative MC showed that SNP-A identified cryptic chromosomal aberrations in 62% and 44% of these patients, respectively.

The clinical relevance of individual new recurrent lesions remains to be established, but it is clear that these lesions contribute to clinical diversity. We analyzed the survival of MDS, MDS/MPD, and sAML patients (irrespective of the type and duration of the therapy received) based on MC and SNP-A findings. We stratified the patients with normal MC into 2 groups: one with a normal karyotype according to both MC and SNP-A, and the other comprised of patients with normal MC in whom new lesions were identified using SNP-A. The latter group showed a reduced overall survival compared with those with normal MC and SNP-A karyotype (median survival 16 vs 39 months, p<.02). We also have studied the impact of SNP-A karyotyping results on survival within MDS, MDS/MPD, and sAML patients treated with diverse regimens separately. While in MDS the survival was not significantly affected, most likely due to short follow-up time (in comparison to the average survival length of MDS patients), in patients with MDS/MPD and sAML newly detected lesions were associated with worse survival. In particular, in sAML newly detected lesions in patients with normal MC conveyed worse prognosis, as did additional lesions in those with already established abnormal MC.

In MDS, we also have tested whether the newly identified lesions have a potential to modify the currently accepted IPSS system (excluding sAML, as this diagnosis is now classified separately, though these patients would have been classified per the French-American-British (FAB) classification as RAEBt and given an IPSS score). MDS patients were grouped only based on their risk category including IPSS-low, Int-1, Int-2, and high-risk groups. For each risk group, patients were subdivided into those in whom SNP-A did not detect additional aberrations and those in whom new lesions were found. The survival curves clearly diverged for patients originally classified as IPSS Int-1, suggesting that SNP-A conveys additional information allowing for better prognostic resolution (median survival 28 vs 9 months, p<.03). Due to a short observation period and variable impact of individual lesions, the survival within the low-risk group did not differ between SNP-A–confirmed karyotypes and those in whom MC was modified.

The work described above has been published in the journal *Blood* (Gondek L, Tiu R, O'Keefe CL, Sekeres MA, Theil KS, Maciejewski JP (2008). Chromosomal lesions and uniparental disomy detected by SNP arrays in MDS, MDS/MPD, and MDS-derived AML. *Blood*. **111**: 1534-1542).