Chromosomal karyotype abnormalities are associated with increased risk of hematologic malignancies

Ursula Martine Schick

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science in Genetic Epidemiology

University of Washington

2012

Committee:

Christopher S. Carlson, PhD

Karen Edwards, PhD

Program Authorized to Offer Degree:
Public Health Genetics

University of Washington

Abstract

Chromosomal karyotype abnormalities are associated with increased risk of hematologic malignancy

Ursula Martine Schick

Chair of the Supervisory Committee:
Affiliate Associate Professor Christopher S. Carlson, PhD
Epidemiology

Chromosomal karyotype abnormalities provide clinical utility in the diagnosis and treatment of hematologic malignancies, and may be predictive of risk of malignant transformation in individuals without apparent clinical presentation of a hematologic malignancy. To assess the association of large-scale chromosomal karyotype abnormalities and hematologic malignancy diagnosed subsequent to specimen collection, we applied the anomDetectBAF algorithm to Genome Wide Association Study data initially conducted using peripheral blood-derived DNA of 9,934 samples from the Women's Health Initiative (WHI). In this sample, large chromosomal karyotype abnormalities were observed at enrollment in 2.39% of the participants, conferring a 2.40-fold increased risk of hematologic malignancy during the median 12.62-year follow-up period (95% CI = 1.22-4.70, p-value= 0.011). Large putatively mosaic chromosomal karyotype anomalies were associated with a 3.11-fold increased risk of a hematologic malignancy in follow-up (95% CI=1.58-6.15, p-value=1.08e-3). This work suggests that large chromosomal karyotype abnormalities detected incidentally in GWAS data may provide clinically relevant risk information for subsequent hematologic malignancy diagnoses in elderly study participants.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

Introduction

Specific Aims

Chromosomal karyotype abnormalities (CKA) provide clinical utility in the diagnosis and treatment of hematologic malignancies, and may be predictive of risk of malignant transformation in individuals without apparent clinical presentation of a hematologic malignancy. In this study, we sought to assess the association between CKA detected at baseline and hematological malignancy (HM) diagnosed in follow-up. Using a large cohort of adults aged 50 years or older, we conducted an observational cohort study with Genome-Wide Association Study (GWAS) data and linked clinical covariates. The cohort for this study was formed with three separate case-control studies of unrelated outcomes from the Women's Health Initiative (WHI). Our specific aims in conducting this study are as follows:

1. Assess the association between CKAs detected at study baseline and subsequent HM diagnosis.
2. Assess the contribution of putative mosaic CKAs detected at study baseline and subsequent HM diagnosis.

Background and Significance

HM are a heterogeneous group of cancers affecting blood and blood-forming tissues (bone marrow and lymphatic system). Deriving from either myeloid or lymphoid cell lines, this category of malignancy includes leukemia, lymphoma and myeloma. Acute and chronic myelogenous leukemia, myelodysplastic syndromes and myeloproliferative disease originate from myeloid cell lines, whereas lymphoma, lymphocytic leukemia and myeloma originate from the lymphoid blood cell lines.

In 2007, in excess of 119,000 HM cases were diagnosed and nearly 55,000 individuals died from HM in the United State alone (U.S. Cancer Statistics Working Group, 2010). Cancers of later adulthood, the incidence of most types of HM increases exponentially with age with a median age of diagnosis between 65-70 years (Lichtman, 2008). It is estimated that HM diagnoses make up approximately 8% of all new cancer diagnoses and constitute 7% of all cancer-related deaths in developed countries (Rodriguez-Abreu, Bordoni, & Zucca, 2007).

On a molecular level, HM frequently display tumor-specific non-random chromosomal karyotypic anomalies (CKA) of both copy number and cytogenetic structure (Yaghmaie et al., 2009). Cytogenetic abnormalities contributing to hematologic and other malignant transformation arise primarily through the aggregation of somatic mutations in sporadic cases, however in familial cancers, predispositional germ-line mutations facilitate transformation (Colotta, Allavena, Sica, Garlanda, & Mantovani, 2009; Croce, 2008; Hanahan & Weinberg, 2000; Hanahan & Weinberg, 2011). Germ-line mutations are thought to be causative in a very small proportion of HM with most HM resulting from successive somatic insults (Segel & Lichtman, 2004). Established non-random tumor-specific CKA have clinical importance as these lesions are important markers for diagnostic, prognostic, and therapeutic modalities in most types of hematological malignancy (Gondek et al., 2008; Tiu et al., 2009; Tiu et al., 2011).

In recent years, many approaches to detecting CKA have been developed to screen for anomalies in GWAS data. These efficient and cost-effective approaches have many benefits over more traditional karyotyping methods, including increased resolution for unbalanced chromosomal alterations and the ability to capture copy-neutral loss of heterozygosity (LeBron et al., 2011; Staaf et al., 2008; Wiech et al., 2009). In addition, development of newer single nucleotide polymorphism (SNP) arrays that contain both SNP and copy number variant probes have led to better resolution of unbalanced chromosomal abnormalities.

Somatic anomalies, such as those frequently identified in HMs, tend to be large (Frohling & Dohner, 2008; Laurie et al., 2012), however no previous studies have investigated anomaly size as a predictor of HM risk. Predicated on the hypothesis that large CKAs detected in GWAS data are likely to represent somatic changes, and therefore correlate with incident HM, our aim was to assess the association between baseline CKA and subsequent HM diagnosis. To our knowledge, no previous studies have investigated CKA size as a predictor of risk of diagnosis HM.

Materials and Methods

The WHI Study Population

The Women's Health Initiative (WHI) is a large cohort study with a primary focus on cardiovascular disease and breast cancer, but with secondary outcomes that include several adjudicated HM (Curb et al., 2003). WHI participants were female, aged 50-79 years at

enrollment, consented for research on age-related health issues, with an average of 12 years follow-up after recruitment. WHI participants with Illumina GWAS data were available from three separate case-control GWAS studies within WHI: a study of colorectal cancer (GECCO; Peters, PI), a study of hip fractures (Hip Fracture; Jackson, PI), and a study of hormone treatment and cardiovascular disease/metabolic outcomes (GARNET; Reiner, PI). WHI approval was obtained for the inclusion of these samples in this study. In total, 9,934 previously genotyped WHI participants were included in the present analysis. We observed no correlation between case status and either detected anomalies or subsequent HM, and, therefore, included all eligible cases and controls in our analyses (**Table 1**).

Classification of Clinical Outcomes of Interest and Relevant Covariates

The primary clinical outcomes of interest in this study were hematological malignancies, including leukemia, lymphoma and multiple myeloma. Ascertainment of clinical outcomes occurred according to previously described WHI methods (Curb et al., 2003). Briefly, HM outcomes were ascertained by self-report at annual or bi-annual follow-up with participants, with subsequent physician adjudication based on medical records and pathology reports to confirm HM cases and assign an International Classification of Disease for Oncology, 3rd edition code when appropriate. With the exception of multiple myeloma, self-reported history of HM prior to WHI enrollment was available for participants. WHI participants with either a reported history of HM at enrollment (n=33) or adjudicated HM diagnosis of HM within six months of enrollment (n=4) were excluded from the present analysis, leaving 9,934 putatively HM-free individuals eligible for the analysis.

All participants were considered to be putatively HM-free at baseline, with neither reported history of an HM, nor an HM diagnosis within the first 6 months of study follow-up. Participants reporting history of HM at baseline or diagnosed with HM in the first 6 months of study follow-up were excluded from all analyses and are not included in subject counts.

Genotyping and Quality Control

WHI samples were genotyped on the following Illumina platforms: HumanOmni1_Quad_v1_0_B, Human 610-Quad, Human HapMap 550K and HUMAN CYTOSNP-12 at the Broad Institute of MIT and Harvard, and the Translational Genomics

Research Institute. Standard quality control was conducted to eliminate samples of unsure identity or DNA quality. We screened for sex discordant samples using estimates of X chromosome heterozygosity, and unintentional duplicates through a global estimate of the proportion identity by state (Purcell et al., 2007). We excluded samples from analysis if genotype call rates were less than 98% or if the B-Allele Frequency (BAF) standard deviation of non-anomalous autosomal chromosomes exceeded 0.06. BAF is a measure of the relative quantity of one allele compared to the other. In total, 522 samples were excluded from the analysis. Of the excluded samples, 219 samples were excluded for high non-anomalous BAF standard deviation of autosomal chromosomes, 298 for low genotyping call rate, and 5 for gender-mismatch. Again, these individuals are not included in the subject counts.

Anomaly Detection

Chromosomal anomalies were detected through the "anomDetectBAF" method from the GWASTools package available for R version 2.14 through the Bioconductor repository (Gogarten, 2011). This method is capable of detecting copy gain and loss for segments of greater than 50 KB in Illumina high-density SNP genotyping data, as well as mosaic copy-neutral loss of heterozygosity. In brief, the "anomDetectBAF" detection method relies on circular binary segmentation (Olshen, Venkatraman, Lucito, & Wigler, 2004) to segment chromosomes based on change-points in the relative allelic intensity (B-allele frequency; BAF). On each chromosome, heterozygous and missing SNPs genotypes are identified, and the BAF at these loci is transformed (tBAF: sqrt(min(BAF,1-BAF,abs(BAF-median BAF))). Anomalous segments are called based on deviation from non-anomalous baseline. The analyses were restricted to autosomal anomalies. Although anomaly detection used only BAF, to classify anomalies into gain/loss/copy-neutral we compared the Log R Ratio (LRR) for the anomalous region to the mean across the genome in an individual. We excluded the sex chromosomes from this LRR metric due to the inherent copy-number difference between males and females(Laurie et al., 2012). Examples of loss and gain anomies are displayed in **Figure 1a** and **1b**, respectively.

We screened for CKAs that were likely due to genotype platform-specific artifacts by binning CKAs into groups by their size and location, then calculating the coverage percentage (percent of samples that had an anomaly in the bin). Bins present only on a single platform or in a single cohort were flagged, and anomalies in that bin were removed from analysis. In total, 7 bins

containing 777 anomalies were identified and removed (8.79%). The largest such bin contained 214 anomalies present on chromosome 8 (position 39497557- 39356825) in only the hip fracture sample processed on the Human 610-Quad platform.

Exposure Classification

Exposures of interest included small (<2 Mb) and large (>=2Mb) CKA stratified by all detected anomalies and mosaic (acquired somatic) anomalies. In total, we considered four exposure groups: all small detected CKA, all large detected CKA, mosaic small detected CKA, and mosaic large detected CKA. All CKA were detected from GWAS data using the anomDetectBAF algorithm (Gogarten, 2011; Laurie et al., 2012).

Statistical Analysis

All statistical analyses were carried out in R version 2.14.0 (R Development Core Team, 2011). We utilized the "coxph" function with the "survival" library for Cox proportional hazard ratio estimates and to fit Kaplan Meier curves, and the "kmeans" function of the "stats" library to cluster putative constitutive anomalies. The "anomStatsPlot" plotting function from the GWASTools package (Gogarten, 2011) was used to generate anomaly figures (**Figure 1a, Figure 1b)**.

We evaluated the association between all anomalies of small (<2Mb) and large (>2Mb) size with incident hematological cancer as unadjusted estimates and after adjusting for age at specimen collection (or study intake) and initial study through modeling of Cox proportional Hazard Ratios, 95% Confidence Intervals and p-values. For all Cox proportional hazard ratio estimates, we used the robust variance option. Survival time was assessed as time between specimen collection/intake and of HM diagnosis for individuals with cancer and as time between specimen collection/intake and death or study attrition for unaffected individuals. Incident HM diagnosis was considered as the event indicator. To ensure that our estimates were robust to the influence of other confounding factors, we performed multivariate analysis including age, sex, ethnicity and initial study as covariates. The magnitudes of risk estimates were largely unchanged by correction for covariates (**Table 3**). We adjusted for age based on the observation that frequency of anomalies generally increases with age (Laurie et al., 2012). Age is also correlated with frequency of HM (**Table 4**) with the median age of onset between 65-70 years (Lichtman, 2008).

Estimates were also adjusted for study. The reference group for the analysis of small anomalies was composed of samples without a detected anomaly. For the analysis of large anomalies, the referent group was individuals with either no detected anomalies or only small anomalies.

We classified CKAs into constitutive (potentially germline copy number variants (CNV)) and mosaic (acquired somatic) based on the observation that anomalies matching previously reported CNV locations primarily fell within the trisomic LRR/BAF region. After plotting Log R Ratio deviation (median(anomalous LRR)- median(nonanomalous LRR)) vs. B-allele frequency deviation (Median|anomalous BAF-Median(nonanomalous BAF)|), we normalized both measures to z-scores and then used k-means clustering to define k=6 clusters of anomalies in the LRR/BAF data. Defining likely constitutive anomalies as those within two median absolute deviations of the median of the cluster at LRR=1.5 and BAF=0.were removed from analyses of putative mosaic anomalies. Clustering of constitutive anomalies is displayed in **Figure 2**. We re-evaluated the risk of HM given a small or large mosaic anomaly after excluding putative constitutive CKAs. Again, risk estimates for mosaic anomalies were adjusted for participant age and initial study. Reference groups were defined consistently with referents in the analysis of all detected anomalies.

Results

Study Characteristics

At least one CKA was detected in 53.8% (n=5,340) of the WHI sample. Most CKAs detected were small (WHI median CKA size=20 kb); large CKAs (≥2 megabases) were detected in approximately 2.4% of samples. The majority of large CKA were predicted to have arisen somatically in WHI (76.8%), whereas the majority of small CKA appear to fit with constitutive mode of inheritance (75.8%). Summary characteristics of large CKA are displayed graphically in **Figure 3.**

Characteristics of putatively HM-free samples with and without detected anomalies were comparable on the basis of age at study intake, race, and duration of study follow-up (**Table 5**). Rates of CKA detection varied by study, which is likely to be an artifact of the density of the genotyping platform utilized in each study and heterogeneity in subject characteristics across studies (**Table 6**). Older age was associated with higher rates of anomaly detection (**Table 4**),

which may partially explain the high anomaly detection among the oldest aged study, hip fracture samples from WHI (median age at intake=70). Frequencies of incident HM in the three WHI studies were comparable to the overall WHI study population (**Table 6**).

Overview of Primary Findings: all anomalies

Among the 9,934 samples in WHI that did not have recorded HM prior to enrollment or within 6 months of enrollment, there were 155 subjects with an adjudicated HM identified during follow-up. Results were consistent with an increased risk of incident HM diagnosis among individuals with CKA greater than 2 megabases (**Table 8**) with a hazard ratio (HR) of 2.40 (95% CI= 1.22-4.70, p=1.11e-02). The association between large CKA and leukemia was pronounced with an estimated HR of 6.51 (95% CI=2.29-18.48, p=4.31e-04) as compared to the referent group. CKA smaller than 2 megabases were not significantly associated with incident HM (**Table 8).** Results were not consistent with a statistically significant difference in probability of remaining HM-free between anomaly classes (**Figure 4**).

Overview of Secondary Findings: mosaic anomalies

In a secondary analysis, we investigated the association between putatively mosaic (acquired somatic) CKA and incident HM. Among participants with large CKA, mosaic anomalies conferred a higher risk of HM than the combined class (**Table 8**). The presence of a large mosaic anomaly was associated with a 3.11-fold increase in risk of HM (95% CI: 1.58-6.15, p=1.08e-03). Small mosaic CKA were observed to be associated with a 1.70-fold excess risk of HM (95% CI: 1.11-2.59, p-value=1.40e-02) relative to individuals without detected CKA. Both small and large mosaic anomalies were strongly associated with a risk of leukemia with HR of 5.23 (95% CI= 2.38-11.46, p=3.36e-05) and 8.71 (95% CI= 3.01-25.22, p=6.58e-05), respectively.

Discussion

Description of Findings

We demonstrate a robust association between CKA of greater than 2 megabases detected in GWAS data and a diagnosis of HM in the following decade. Despite this interesting finding, it is important to consider that these findings are based on a very small case series and are not

replicated at this point. In addition, the study composition of predominantly Caucasian race (>92%) and a bias toward female sex may limit the generalizability of these findings. Further research is warranted to independently confirm these interesting findings in larger, more diverse populations.

To our knowledge, this is the first report of chromosomal anomaly size as a predictor of HM risk. Providing some evidence for our findings for mosaic abnormalities, a recent study by Laurie et al. reports that chromosomal mosaicism was associated with a 10-fold increased risk of incident hematologic cancer (Laurie et al., 2012). In addition, it has been observed that most chromosomal imbalances in cancer tend to be large(Frohling & Dohner, 2008), which supports our finding of the robust association specifically with large anomalies. A previous study investigating 9p deletions in chronic myeloid leukemia patients reported that large (>1.4 Mb), but not small deletions were associated with decreased survival (Fourouclas et al., 2006). These studies provide some basis for comparison of our findings, however further studies are required to substantiate this study.

Study Strengths and Limitations

There were several strengths of the study. A major strength of the study was the availability of paired DNA sample and long-range clinical data on most participants. The availability of clinical data facilitated the detection of HM prior to enrollment as part of the exclusion criteria and the detection of HM outcomes in the extensive follow-up period. Unlike most epidemiologic studies, the active WHI adjudication for outcomes presented the opportunity to assess risk of HM following baseline over a fairly long duration. Detailed reports in the form of medical billing codes (ICD-O3) provided sufficient information in most cases for high confidence in diagnosis.

Another strength of the study was the stability in participation. The WHI cohorts were comprised of very stable populations with low attrition rates. Based on these features, the WHI case-control studies were excellent population sources for this study by providing long and consistent forms of follow-up data.

Despite the many enumerated strengths of the study, there are several notable limitations to the study. Firstly, the use of multiple study populations led to heterogeneity in participant

characteristics and study covariates. The WHI cohorts were all sampled as case-control studies for a given outcome. Outcomes varied across studies and characteristics of each case and control set differed.

In addition to participant characteristic heterogeneity, study design features including, but not limited to, genotyping platform, genotyping center, and covariate collection differed across study. All genotypes were collected from Illumina-based platforms, however densities of SNP and intensity only probes varied and protocols were likely to be different. To the extent to which genotyping center and probe density mattered, heterogeneity in this covariate may have also modified study findings.

Lastly, the results from this study should be interpreted cautiously as the estimates are based on relatively small numbers without proper replication. Given the relative rarity of HM outcomes, future cohort studies seeking to further this research should be based on very large sample sizes in an effort to ascertain more cases. Alternatively, case-control studies can be used to ensure adequate HM case numbers.

Implications and Future Directions

Though the association between abnormality size and HM risk is not well characterized, our findings are nevertheless suggestive of excess risk, which might approach the level at which an incidental, yet clinically relevant research result could be considered for return to research participants. Return of incidental research results is a complicated issue. Prior to further discussion of these results as a returnable research result, these results clearly require further, independent confirmation in other populations.

In addition to confirmatory studies, further research of these data should seek to investigate mechanisms underlying the observed association. In particular, there should be careful characterization of the CKA to detect enrichment in genes known to be involved in cancer (tumor suppressor genes, oncogenes, and DNA repair genes) and to clarify the basis of the observed association.

In summary, this work provides the foundation for future research to explore the association between anomaly size and hematologic outcomes. In particular, research utilizing a trio

(offspring, mother, father) design could be utilized to provide stronger evidence for the differences in risk of HM between somatic and constitutive abnormalities.

Acknowledgments

Firstly, I would like to acknowledge the contribution of Christopher S. Carlson, Karen Edwards and Andrew McDavid for their ongoing input, support, time and resources. I wish to thank all of the study collaborators (Paul K. Crane, Noah Weston, Kelly Ehrlich, Katherine M. Newton, Robert Wallace, Ebony Bookman, Tabitha Harrison, Aaron Aragaki, Sophia S. Wang, Alex P. Reiner, Rebecca D. Jackson, Ulrike Peters, Eric B. Larson, and Gail P. Jarvik). I thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whiscience.org/publications/WHI_investigators_shortlist.pdf. WHI GECCO colorectal cancer study was funded by U01CKA137088 and R01CKA059045. I wish to acknowledge the support of the GARNET Collaborative Research Group. Funding support for WHI-GARNET was provided through the NHGRI Genomics and Randomized Trials Network (GARNET) (Grant Number U01 HG005152). Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GARNET Coordinating Center (U01 HG005157). Funding support for genotyping, which was performed at the Broad Institute of MIT and Harvard, was provided by the NIH Genes, Environment and Health Initiative [GEI] (U01 HG004424).

Figures and Tables

Figures

Figure 1. a) Example of a detected loss of genetic material plotted with the "anomStatsPlot" function. **b)** Example of a mosaic gain of chromosome 8 plotted with the "anomStatsPlot" function.

a) **b)**

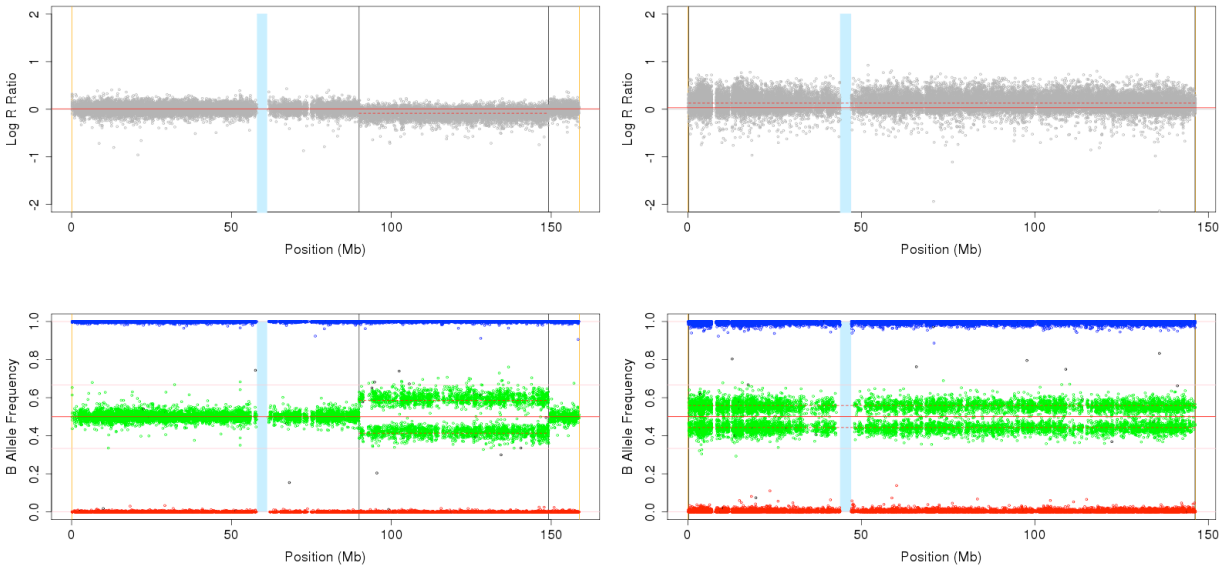


Figure 2. Clustering of putative mosaic, constitutional and whole chromosome anomalies in WHI.

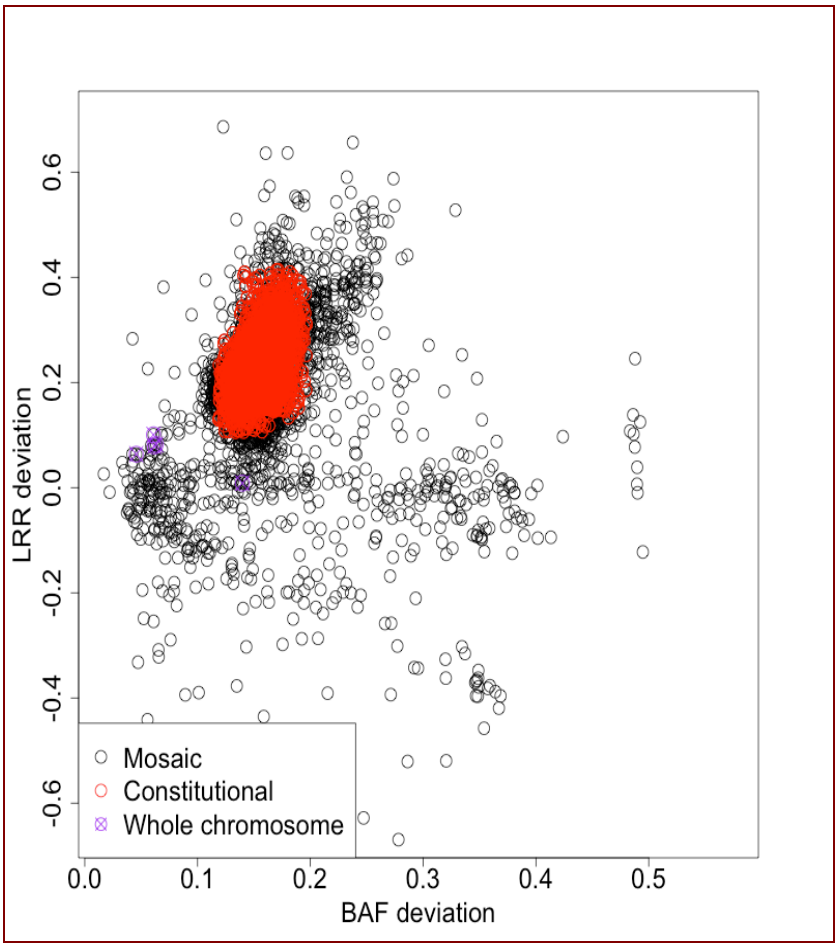
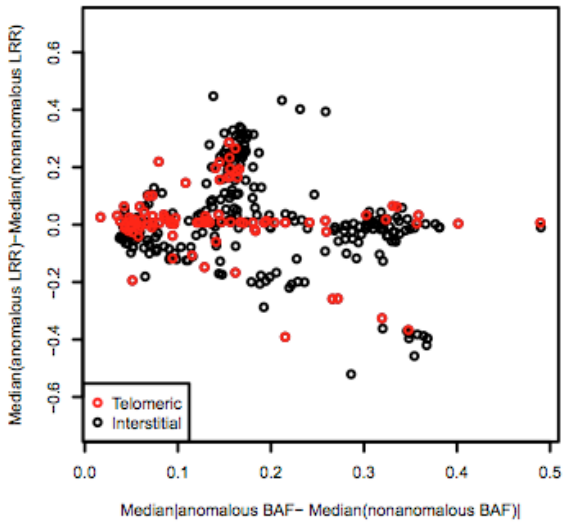
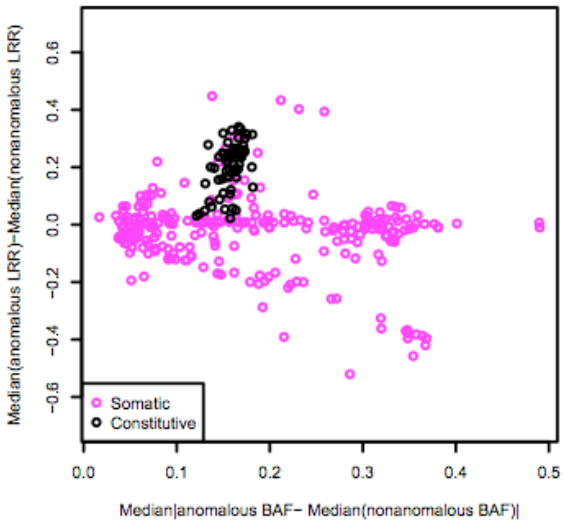


Figure 3. Characteristics of Large Anomalies detected in WHI. **a)** Large CKA in WHI classified by telomeric (red) or interstitial (black). **b)** Large CKA in WHI classified by putative somatic (magenta) or constitutive (black). **c)** Large CKA in WHI classified by copy gain (red), copy neutral (green), copy loss (blue) and ambiguous (neutral/gain, neutral/loss; cyan).

a)



b)



c)

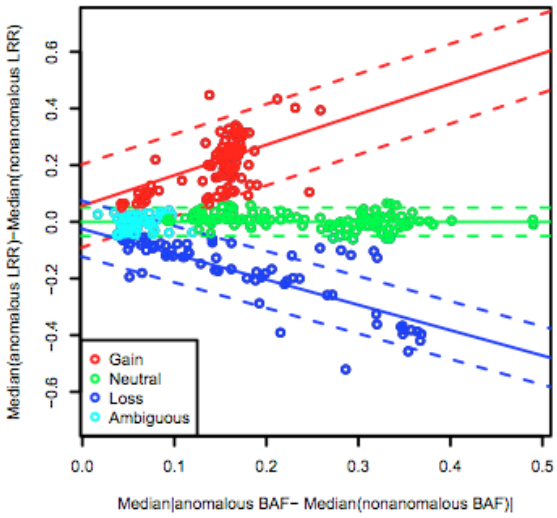
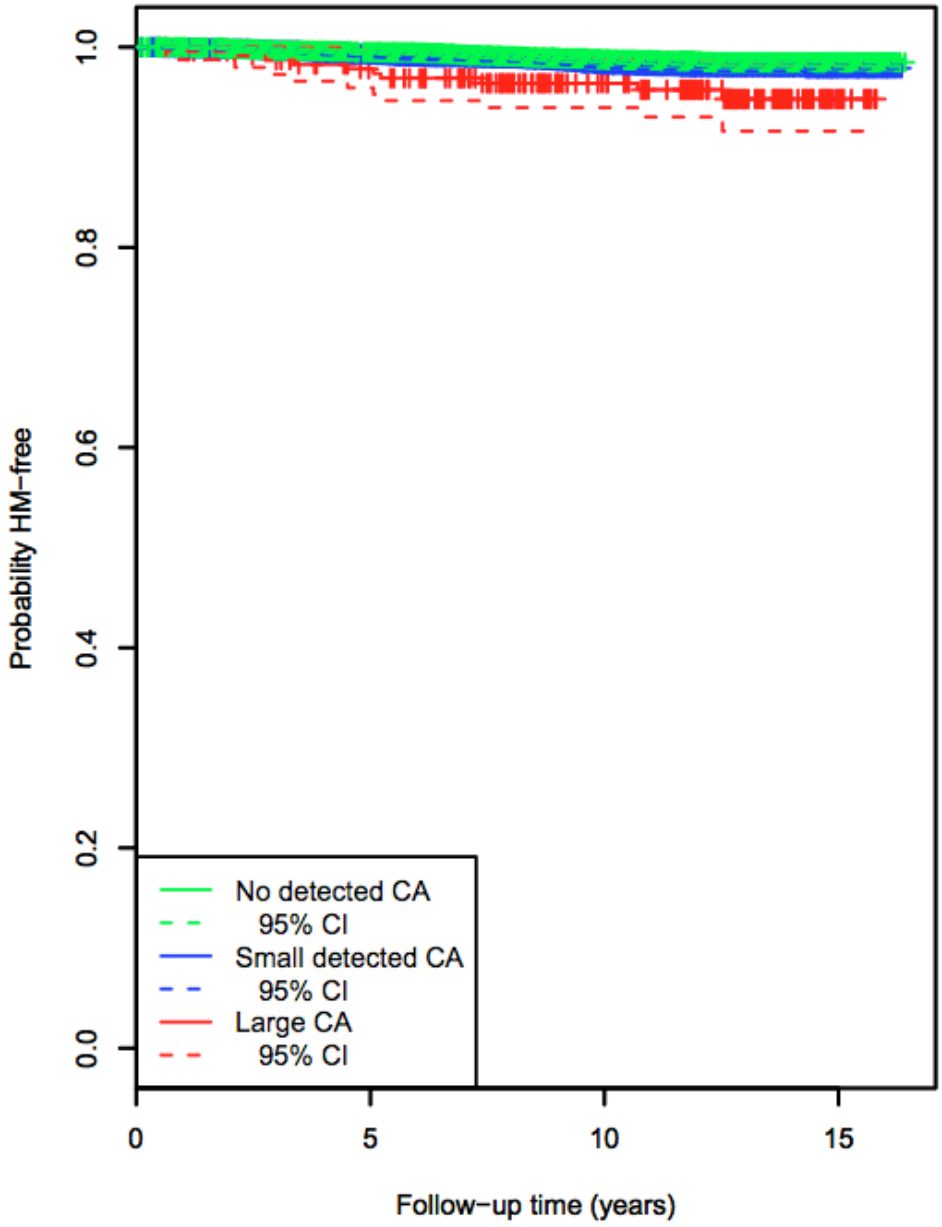


Figure 4. Kaplan-Meier plot modeling the probability of remaining hematologic malignancy-free over follow-up in WHI. Vertical lines represent censoring times.



Tables

Table 1: Pearson correlation coefficients relating case status to both anomaly size and incident HM.

Study: outcome (n)	Case in prior study n(%)	Control in prior study n (%)	Correlation of Case Status and Any CKA r^2[95% CI]	Correlation of Case Status and Large CKA (>2MB) r^2[95% CI]	Correlation of Case Status and small CKA (<2MB) r^2[95% CI]	Correlation of Case Status and Incident HM r^2[95% CI]
WHI GARNET: CV/metabolic (n=4,699)	2,434 (51.80)	2,265 (48.20)	-0.016 [-0.045, 0.012]	0.021 [0.049, 0.008]	-0.011 [-0.039,0.018]	0.010 [-0.018, 0.039]
WHI Hip Fracture (n=3,239)	1,937 (59.80)	1,302 (40.20)	-0.050 [-0.084, -0.016]	0.033 [-0.002, 0.067]	-0.006 [-0.096, -0.028]	-0.003 [-0.047, 0.041]
WHI GECCO: Colorectal Cancer (n=1,996)	987 (49.45)	1009 (50.55)	-0.029 [-0.072, 0.015]	0.038 [-0.006, 0.0817]	-0.038 [-0.082, 0.006]	0.020 [-0.014, 0.054]

Table 2. Classification of HM by ICD-O3 medical billing codes.

Hematologic Diagnosis	WHI (ICD-O3 Histology Code)
Hodgkin Lymphoma	9650, 9663, 9680
Non-Hodgkin Lymphoma	9590, 9591, 9670, 9671, 9673, 9680, 9684, 9687, 9689, 9690, 9691, 9695, 9698, 9699, 9702, 9705, 9714, 9729
Leukemia	9823,9834,9836,9840,9861,9874,9875,9895, 9940,9945
Multiple Myeloma	9731, 9732

Table 3. Adjusted and unadjusted Cox Proportional Hazard Ratios (HR) with associated 95% confidence interval and Bonferroni-corrected p-values by anomaly type and HM.

HM	CKA Type	HR	95% CI L	95% CI U	P-value
All HM in WHI (n=155)	Large (adjusted)	2.4	1.22	4.7	1.11E-02
All HM in WHI (n=155)	Large (unadjusted)	2.61	1.35	5.02	4.22E-03
All HM in WHI (n=155)	Small (adjusted)	1.29	0.9	1.85	1.65E-01
All HM in WHI (n=155)	Small (unadjusted)	1.34	0.96	1.87	8.22E-02
All HM in WHI (n=155)	Large, Mosaic (adjusted)	3.11	1.58	6.15	1.08E-03
All HM in WHI (n=155)	Large, Mosaic (unadjusted)	3.45	1.79	6.66	2.17E-04
All HM in WHI (n=155)	Small, Mosaic (adjusted)	1.7	1.11	2.59	1.40E-02
All HM in WHI (n=155)	Small, Mosaic (unadjusted)	1.78	1.18	2.7	5.63E-03
Leukemia in WHI (n=35)	Large (adjusted)	6.51	2.29	18.48	4.31E-04
Leukemia in WHI (n=35)	Large (unadjusted)	6.21	2.3	16.74	3.09E-04
Leukemia in WHI (n=35)	Small (adjusted)	1.49	0.66	3.33	3.35E-01
Leukemia in WHI (n=35)	Small (unadjusted)	1.38	0.67	2.87	3.86E-01
Leukemia in WHI (n=35)	Large, Mosaic (adjusted)	8.71	3.01	25.22	6.58E-05
Leukemia in WHI (n=35)	Large, Mosaic (unadjusted)	8.23	3.05	22.22	3.22E-05
Leukemia in WHI (n=35)	Small, Mosaic (adjusted)	5.23	2.38	11.46	3.63E-05
Leukemia in WHI (n=35)	Small, Mosaic (unadjusted)	4.73	2.28	9.8	2.96E-05
Multiple Myeloma in WHI (n=30)	Small (adjusted)	2.43	1.08	5.47	3.21E-02
Multiple Myeloma in WHI (n=30)	Small (unadjusted)	2.53	1.12	5.69	2.50E-02
Multiple Myeloma in WHI (n=30)	Small, Mosaic (adjusted)	2.54	1.31	5.69	2.37E-02
Multiple Myeloma in WHI (n=30)	Small, Mosaic (unadjusted)	2.57	1.15	5.78	2.14E-02
Non-Hodgkin Lymphoma in WHI (n=85)	Large (adjusted)	2.21	0.87	5.66	9.73E-02
Non-Hodgkin Lymphoma in WHI (n=85)	Large (unadjusted)	2.5	0.99	6.28	5.21E-02
Non-Hodgkin Lymphoma in WHI (n=85)	Small (adjusted)	1.09	0.66	1.79	7.36E-01
Non-Hodgkin Lymphoma in WHI (n=85)	Small (unadjusted)	1.18	0.76	1.84	4.55E-01
Non-Hodgkin Lymphoma in WHI (n=85)	Large, Mosaic (adjusted)	2.85	1.11	7.33	2.99E-02
Non-Hodgkin Lymphoma in WHI (n=85)	Large, Mosaic (unadjusted)	3.31	1.32	8.33	1.10E-02
Non-Hodgkin Lymphoma in WHI (n=85)	Small, Mosaic (adjusted)	0.7	0.34	1.47	3.49E-01
Non-Hodgkin Lymphoma in WHI (n=85)	Small, Mosaic (unadjusted)	0.79	0.38	1.64	5.29E-01

Table 4: Number and frequency of incident hematologic outcomes and CKA by classification in WHI.

Age in years (n)	Incident HM n (%)	Small CKA detected n (%)	Small mosaic CKA detected n (%)	Large CKA detected n (%)	Large mosaic CKA detected n (%)
All Ages (n=9934)	155 (1.56)	5103 (51.37)	1,233 (12.41)	237 (2.39)	182 (1.83)
50-54 (n=549)	0 (0.00)	283 (51.55)	67 (12.20)	10 (1.82)	7 (1.28)
55-59 (n=1199)	8 (0.67)	591(49.29)	125 (10.43)	14 (1.17)	8 (0.67)
60-64 (n=1897)	27 (1.42)	942 (49.66)	215 (11.33)	36 (1.90)	28 (1.48)
65-69 (n=2548)	57 (2.24)	1298 (50.94)	315 (12.36)	47 (1.84)	31 (1.22)
70-74 (n=2390)	40 (1.67)	1251 (52.34)	326 (13.64)	75 (3.14)	63 (2.64)
74-79 (n=1351)	23 (1.70)	738 (54.63)	185 (13.69)	55 (4.07)	45 (3.33)

Table 5. Characteristics and HM outcomes by anomaly status in WHI.

Characteristics	No CKA (n=4,594)	All Small CKA (n=5,103)	Small Mosaic CKA (n=1,233)	All Large CKA (n=237)	Large Mosaic CKA (n=182)
Age (yr) at Entry mean(sd)	66.23 (6.85)	66.76 (7.00)	67.15 (6.94)	68.80 (6.65)	69.35 (6.50)
Sex (% Female)	100	100	100	100	100
Race (% Caucasian)	92.53	92.73	92.05	91.14	90.66
Median follow-up (yr)	12.66	12.58	12.34	11.94	11.84
Clinical Outcomes					
All HM (n)	59	86	29	10	10
Hodgkin Lymphoma (n)	4	1	1	0	0
Non-Hodgkin Lymphoma (n)	35	45	8	5	5
Leukemia (n)	12	18	12	5	5
Multiple Myeloma (n)	8	22	8	0	0

Table 6: Anomaly and incident HM detection rates by initial study group.

Study (n)	**Participants with Detected Anomaly n (%)**	**Individuals with any qualifying hematologic outcomes cases(n) (% of study)**	**Individuals with Incident Hodgkin Lymphoma cases(n) (% of study)**	**Individuals with Incident Non-Hodgkin Lymphoma cases(n) (% in study)**	**Individuals with Incident Leukemia cases n (% in study)**	**Individuals with Incident Multiple myeloma cases n (% in study)**
All Subjects (n=9934)	5340 (53.75)	155 (1.56)	5 (0.05)	85 (0.86)	35 (0.35)	30 (0.30)
GARNET (n=4699)	2576 (54.82)	68 (1.45)	2 (0.04)	30 (0.64)	18 (0.38)	18 (0.38)
GECCO (n=1996)	577 (28.91)	32 (1.60)	1 (0.05)	20 (1.00)	8 (0.40)	3 (0.15)
Hip Fractures (n=3239)	2187 (67.52)	55 (1.70)	2 (0.06)	35 (1.08)	9 (0.28)	9 (0.28)

Table 7. Characteristics of the WHI study population stratified by HM status.

	All subjects putatively HM-free at baseline (N=9,934)**	No Detected HM in follow-up (n=9,779)	Incident HM cases (n=155)
Years of Age at Study Enrollment mean(sd)	66.56 (6.94)	66.53 (6.96)	68.46 (5.13)
Sex (% Female)	100	100	100
Race (% Caucasian)	92.62	92.57	96.13
Median follow-up (years)	12.62	12.62	11.78
Median time from specimen collection* to diagnosis (years)	—	—	6.15

*Study intake was used as a proxy for DNA collection if date of DNA collection omitted from individual characteristics.

Table 8. Adjusted Cox proportional hazard ratio estimates assessing the risk of incident HM associated with chromosomal anomalies stratified by size and mosaic status.

	Small CKA (<2 Mb) Referent: no detected anomalies		**Large CKA (>2 Mb) Referent: no detected or small anomalies**	
Incident Hematologic Outcome	**HR (all CKA) [95% CI, p-value]**	**HR (Mosaic CKA) [95% CI, p-value]**	**HR (all CKA) [95% CI, p-value]**	**HR (Mosaic CKA) [95% CI, p-value]**
Hodgkin Lymphoma (n=5)	--	--	--	--
Non-Hodgkin Lymphoma (n=85)	1.09 [0.66-1.79, 7.36e-01]	0.70 [0.34-1.47, 3.49e-01]	2.21 [0.87-5.66, 9.73e-02]	2.85 [1.11-7.33, 2.99e-02]
Leukemia (n=35)	1.49 [0.66-3.33, 3.35e-01]	5.23 [2.38-11.46, 3.63e-05]	6.51 [2.29-18.48, 4.31e-04]	8.71 [3.01-25.22, 6.58e-05]
Multiple Myeloma (n=30)	2.43 [1.08-5.47, 3.21e-02]	2.54 [1.31-5.69, 2.37e-02]	--	--
All Hematologic Malignancies (n=155)	1.29 [0.90-1.85, 1.65e-01]	1.70 [1.11-2.59, 1.40e-02]	2.40 [1.22-4.70, 1.11e-02]	3.11 [1.58-6.15, 1.08e-03]

* WHI estimates are adjusted for age at intake and initial study group.

References

Colotta, F., Allavena, P., Sica, A., Garlanda, C., & Mantovani, A. (2009). Cancer-related inflammation, the seventh hallmark of cancer: Links to genetic instability. *Carcinogenesis, 30*(7), 1073-1081. doi:10.1093/carcin/bgp127

Croce, C. M. (2008). Oncogenes and cancer. *The New England Journal of Medicine, 358*(5), 502-511. doi:10.1056/NEJMra072367

Curb, J. D., McTiernan, A., Heckbert, S. R., Kooperberg, C., Stanford, J., Nevitt, M., . . . WHI Morbidity and Mortality Committee. (2003). Outcomes ascertainment and adjudication methods in the women's health initiative. *Annals of Epidemiology, 13*(9 Suppl), S122-8.

Fourouclas, N., Campbell, P. J., Bench, A. J., Swanton, S., Baxter, E. J., Huntly, B. J., & Green, A. R. (2006). Size matters: The prognostic implications of large and small deletions of the derivative 9 chromosome in chronic myeloid leukemia. *Haematologica, 91*(7), 952-955.

Frohling, S., & Dohner, H. (2008). Chromosomal abnormalities in cancer. *The New England Journal of Medicine, 359*(7), 722-734. doi:10.1056/NEJMra0803109

Gogarten, S. M. (2011). In Laurie C., . . . Swarnkar R.(Eds.), *Classes for storing very large GWAS data sets and annotation, and functions for GWAS data cleaning and analysis.* (1.19th ed.) Bioconductor.

Gondek, L. P., Tiu, R., O'Keefe, C. L., Sekeres, M. A., Theil, K. S., & Maciejewski, J. P. (2008). Chromosomal lesions and uniparental disomy detected by SNP arrays in MDS, MDS/MPD, and MDS-derived AML. *Blood, 111*(3), 1534-1542. doi:10.1182/blood-2007-05-092304

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell, 100*(1), 57-70.

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell, 144*(5), 646-674. doi:10.1016/j.cell.2011.02.013

Laurie, C., Laurie, C., Rice, K., Doheny, K. F., Zelnick, L. R., McHugh, C., . . . et al. (2012). A population survey of somatic mosaicism for large chromosomal anomies and its relationship to ageing and cancer. *Nature Genetics, (In Press)*

LeBron, C., Pal, P., Brait, M., Dasgupta, S., Guerrero-Preston, R., Looijenga, L. H., . . . Hoque, M. O. (2011). Genome-wide analysis of genetic alterations in testicular primary seminoma using high resolution single nucleotide polymorphism arrays. *Genomics, 97*(6), 341-349. doi:10.1016/j.ygeno.2011.02.011; 10.1016/j.ygeno.2011.02.011

Lichtman, M. A. (2008). Battling the hematological malignancies: The 200 years' war. *The Oncologist, 13*(2), 126-138. doi:10.1634/theoncologist.2007-0228

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics, 81*(3), 559-575. doi:10.1086/519795

R Development Core Team. (2011). *R: A language and environment for statistical computing* (2.14.0 ed.). Vienna, Austria: R Foundation for Statistical Computing.

Rodriguez-Abreu, D., Bordoni, A., & Zucca, E. (2007). Epidemiology of hematological malignancies. *Annals of Oncology : Official Journal of the European Society for Medical Oncology / ESMO, 18 Suppl 1*, i3-i8. doi:10.1093/annonc/mdl443

Segel, G. B., & Lichtman, M. A. (2004). Familial (inherited) leukemia, lymphoma, and myeloma: An overview. *Blood Cells, Molecules & Diseases, 32*(1), 246-261.

Staaf, J., Lindgren, D., Vallon-Christersson, J., Isaksson, A., Goransson, H., Juliusson, G., . . . Ringner, M. (2008). Segmentation-based detection of allelic imbalance and loss-of-heterozygosity in cancer cells using whole genome SNP arrays. *Genome Biology, 9*(9), R136. doi:10.1186/gb-2008-9-9-r136

Tiu, R. V., Gondek, L. P., O'Keefe, C. L., Elson, P., Huh, J., Mohamedali, A., . . . Maciejewski, J. P. (2011). Prognostic impact of SNP array karyotyping in myelodysplastic syndromes and related myeloid malignancies. *Blood, 117*(17), 4552-4560. doi:10.1182/blood-2010-07-295857

Tiu, R. V., Gondek, L. P., O'Keefe, C. L., Huh, J., Sekeres, M. A., Elson, P., . . . Maciejewski, J. P. (2009). New lesions detected by single nucleotide polymorphism array-based chromosomal analysis have important clinical impact in acute myeloid leukemia. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology, 27*(31), 5219-5226. doi:10.1200/JCO.2009.21.9840

U.S Cancer Statistics Working Group. United States Cancer Statistics: 1999-2007 Incidence and Mortality Web-based Report. Atlanta (GA): Department of Health and Human Services, Center for Disease Controls and Prevention, and National Cancer Institute; 2010. Available at: http://www.cdc.gov/uscs.

Wiech, T., Nikolopoulos, E., Weis, R., Langer, R., Bartholome, K., Timmer, J., . . . Werner, M. (2009). Genome-wide analysis of genetic alterations in barrett's adenocarcinoma using single nucleotide polymorphism arrays. *Laboratory Investigation; a Journal of Technical Methods and Pathology, 89*(4), 385-397. doi:10.1038/labinvest.2008.67

Yaghmaie, M., Gerayeli, N., Ghaffari, S., Ghavamzadeh, A., & Tootian, S. (2009). Some specific chromosomal aberrations of hematologic malignancies in 80 iranian population. *International Journal of Hematology-Oncology and Stem Cell Research, 3*, 28-33.