Single Nucleotide Polymorphisms and Incidence of Lymphoma

Lucy Brining

Abstract Lymphoma is the fifth most commonly diagnosed cancer in the United States and incidence levels are increasing with little new information on possible causes. Several cancers are known to be associated with genetic aberrations. This study investigates the relationship between developing lymphoma and genetic aberrations known as single nucleotide polymorphisms (SNPs), found in the MIF, CY1PA, APM1, and LEP genes. SNPs are single nucleotide base changes within a gene, whose presence is posited to interfere with the genes function. The study tested whether the presence of one or more of the SNPs in question and the interactions between two SNPs increases the risk of developing lymphoma. To investigate the hypothesis, DNA was isolated from 1,236 controls and subjects, with the subjects being cancer patients and analyzed for presence of the four SNPs using Real-Time PCR analysis. The data were tested for Hardy-Weinburg equilibrium to determine normalcy of the data set and analyzed using chi-squared analyses looking for associations between the presence of the SNPs and lymphoma. Results are expected to show that the presence of two of the SNPs will have synergistic interactions and are associated with the development of lymphoma. In addition, the results might show the presence of at least one of the SNPs correlates with the development of lymphoma. Findings from this study could serve as potential biomarkers for cancer in the future.

Introduction

With only a five year survival rate of 52% in the US, lymphomas are the fifth most common cancer affecting about 60,900 people a year (Chang *et al.* 2003). Despite diagnosis improvements and considerable efforts to identify possible risk factors, the causes of most lymphoma cases are unknown and the incidence of lymphoma is increasing (Chang *et al.* 2003). Since 1970, the incidence of lymphoma has nearly doubled (American Cancer Society, 2002). With rates of lymphoma increasing and little understanding why, there is a need for studies investigating possible factors.

Previous studies have looked at associations between lymphoma and environmental and occupational exposures to chemicals including pesticides. Few of these studies however, resulted in consistent conclusions and in fact many contradicted one another (Zheeb and Blettner 1998). Medically-applied radio and chemotherapy have been reported as high-risk factors, however again inconsistently. Hereditary factors like chromosomal syndromes and genetically inherited diseases are proven to put an individual in a high-risk category; however, they are extremely rare in the population (Zheeb and Blettner 1998). Even though such studies correlate the development of the disease with certain factors, few of these studies draw solid conclusions about the known risks, for none of the risks are absolute. Conclusions drawn from such studies are inconsistent with one another due to the fact that cancers are not caused by exposure to one chemical or radiation treatment but rather a combination of events.

In an attempt to investigate synergistic relationships between risk factors like smoking, drinking or diet, there have been case studies looking at whether the incidence of lymphomas is affected by such risk factors (Zheeb and Blettner 1998). However, there have been very few studies investigating synergistic relationships between genes. Thus, there have been few attempts at identifying groups of genes which could serve as potential biomarkers for lymphoma. The few that have been determined are genes that lead to breast cancer, pancreatic cancer and gastrointestinal cancer (Mimori *et al* 02, Kawakam *et al* 99, Chang *et al* 03).

The aim of this study is to identify possible new biomarkers, so that in the future, populations can be tested to evaluate the risk of developing lymphoma. In order to develop such biomarkers, we need to investigate possible genes and the synergistic

relationships between genes. For this study we are looking specifically at single nucleotide polymorphisms or SNPs. SNPs are single nucleotide-base changes in a genetic sequence which on their own do not result in a change of phenotype. The SNP could, in theory, go unnoticed for an entire lifetime. However, it is posited that such base changes in combination with other factors could become the starting points for cancer development. For instance, if the SNP codes for a different amino acid within the protein and if there is some change in the gene environment, the amino acid coded by the SNP might act differently, changing the shape and thus the function of the protein (Skibola *et al* 2002). The change in function could lead to over or under-expression of the gene, resulting in the development of cancer.

The presence of SNPs is not significant for all genes, only genes whose over or under-expression could result in cancerous growth. In our case, the specific genes in question for this study are: MIF, CYP1A, LEP, and AMP1. An interesting characteristic of the MIF gene is that it is thought to control cell proliferation, cell survival, angiogenesis, cell differentiation, T Lymphocytic activation and thus tumor progression (Chesney *et al*. 1998, Takahashi *et al* 1998, Wistow *et al* 1993). There are also reports that MIF is over-expressed in many cancers including leukemia (Chesney *et al*. 1998, Huggins and Fukunishi 1964). CYP1A encodes an enzyme that converts environmental pro-carcinogens to reactive intermediates with carcinogenic effects. Additionally, studies have shown that CYP1A catalyzes the oxidative metabolites of estrogens, another possible starting point for cancer (Chang *et al* 2003). Studies found that SNPs in the CYP1A gene are associated with increased cancer risks (Chang *et al* 2003). LEP is ubiquitous in recent literature correlating the gene with obesity. However, in addition to obesity, LEP also controls hematopoiesis, angiogenesis and immune and inflammatory responses. LEP is also found to be involved in certain cancers. In their 2003 study, Comings *et al* found that the LEP gene is related to breast cancer. Finally, AMP1 is the adipose tissue's most abundant gene transcript and is exclusively expressed in adipose tissue. It is similar to the LEP gene in that it is thought to have some correlation with obesity. Diez and Iglesias found in March 2003 that APM1 also decreases inflammatory responses. All of the above genes serve as possible starting points for cancer and thus are good areas to study.

For the study, we collected blood samples from San Francisco a population. From the blood samples, we extracted DNA and performed Real-Time PCR analysis to evaluate the DNA for the presence or absence of the SNP in question. The goal of the study is to discover whether a SNP's presence in MIF, CYP1A, LEP and APM1 increases the risk of developing lymphoma.

Methods

A population-based case-control study was conducted looking at a San Francisco population with 1256 cases and controls. The study subjects (16-70 years old), included people who had lymphoma that was not induced by chemotherapy and control subjects who had no prior history of lymphoma or any other cancer. The controls and test subject were approached with their physician's permission. The controls were around the same age as the case subjects (+/- 2 years, similar age and race) and there were approximately two controls per study subject.

I extracted the genomic DNA from the peripheral blood by using the QIAmp Blood mini Kit (Qiagen Inc., Valencia, CA) and followed the instructions of the manufacturer to accurately extract the DNA. Once the DNA was extracted, I diluted it to a 10 μg/ml solution with dH_2O.

For the PCR reaction, 1μl of the DNA dilution was then placed into all wells in a 96 welled plate (Applied Biosystems, Foster City, CA) along with one of the following primers used to amplify the loci of the genes of interest. For APM1 I used the forward primer with the following sequence:

5′-GGCCTCTTTCATCACAGACC-3′and the reverse primer with the following sequence: 5′-ACTTTGGCTTTGCTGCATCT-3′. The CYP1A forward primer had the following sequence: 5′-TCACTCGTCTAAATACTCACCCTG-3′ and the reverse primer's sequence was 5′-TAGGAGTCTTGTCTCATGCCT-3′. For MIF I used the forward primer with the sequence of 5′-CGATTTCTAGCCGCCAAGTG-3′ and the reverse primer with the sequence 5′-AGCAACCGCCCGCTAAGC-3′. The primer' sequence for LEP was 5′-GGAGCCCCGTAGGAATCG-3′ and the reverse primer had the following sequence: 5′-TCCTTCCTCCTTCTCTGCTGG-3′ (Applied Biosystems, Foster City, CA). In order to discover whether the SNP in question was present, I used a

fluorogenic 3′ minor groove binding probe in a real-time PCR assay (Kutyavin *et al* 2000). Each well contained approximately 15 µl, which contained 200nm of each probe, 900 nm of forward primer and reverse primer, 10ng of DNA and 9X Taqman Universal PCR Master Mix (Applied Biosystems, Foster City, CA). The PCR was conducted and analyzed in an ABI Prism 7700 thermocycler (Applied Biosystems, Foster City, CA). The PCR thermal cycling conditions consisted of one 2 minute cycle at 50°C, followed by one 10 minute cycle at 95°C cycle, followed by thirty-eight cycles at 92°C, and finishing with one 1 minute cycle at 62°C. The computer followed the progression of the sample by measuring the degree of fluorescence and the degree of fluorescence identifies whether the SNP is present (Kutyavin *et al* 2000).

The distribution of the different genotypes were determined and compared. I described the data by using simple arithmetic means, looking at both the controls and the subjects. I analyzed correlation between polymorphisms and lymphoma by using the chi-squared test with one degree of freedom (DF) and a 95% confidence interval (CI). We also investigated correlations between more than one SNP and the incidence of cancer, again using chi-squared test with 1 DF and a 95% CI.

Results

Due to difficulties regarding access to the complete data set, the analysis and the results of the study are not complete and will not be complete until at least the end of the year. However, the distribution of the various genotypes is known. Out of the total population, LEP showed the highest number of people homozygous for the SNPs in question (Table 1). Only 26 % were homozygous for the CYP1A SNP and even fewer of the study subjects were homozygous for the remaining two SNPs, MIF and APM1.

	CY1PA	APM1	LEP	MIF
1=wt/wt	307	657	172	824
2=wt/var	642	490	530	377
3=var/var	279	82	550	43
0=failed	8	8	10	10
TOTAL	1236	1237	1262	1254

Table 1. Summarizes the results of the actual distribution from the incomplete but real data set. The numbers represent number of people. The number allows us to derive a general idea of what the frequency

of the SNPs is. People who are homozygous for the variant (var) express the SNP, while those who are homozygous for the wild-type (wt) or heterozygous do not express the SNP.

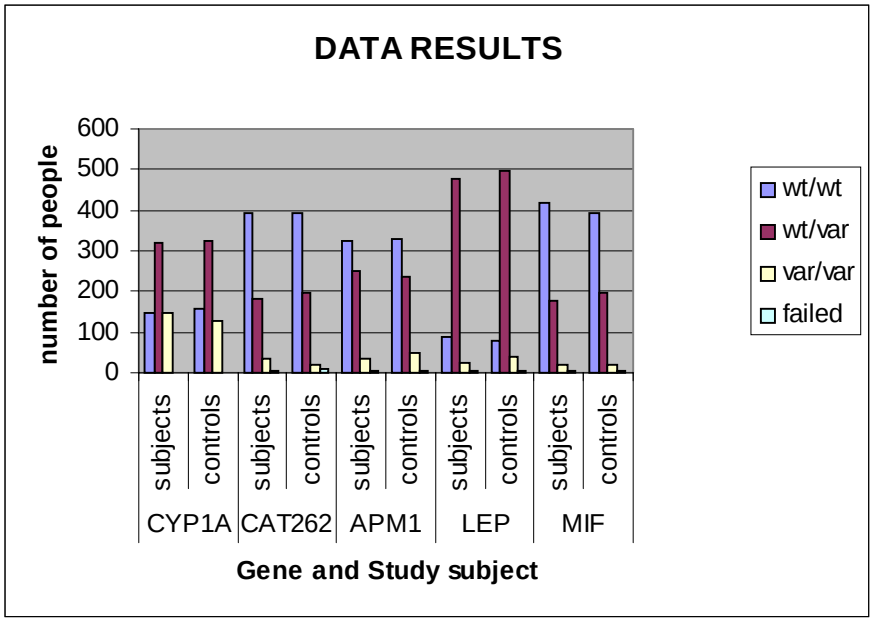
In place of the complete data set, I used a randomly generated data set. The results of these are depicted in Table 2 and graph 1. The results found all of the polymorphisms by themselves not statistically significant (95% CI and 1 DF; see Table 4). This is not too surprising considering all data were randomly generated. The presence of two single nucleotide polymorphisms also was determined to be not statistically significant (CI of 95% and 1 DF; see Table 4).

When the actual analysis is completed, there are two possible outcomes. The first possibility is that there is no correlation between the incidence of lymphoma and SNPs. No correlation would have the same results derived from the random data set, that is, a chi-squared value less than the critical value using 1 DF and 9% CI. The second possibility is that there is a correlation between SNPs and incidence of lymphoma. A correlation would be demonstrated by a chi-squared value greater than the critical value, again using 1 DF and 95% CI.

	1	**2**	**3**	**4**
CY1PA				
subjects	1 48	3 19	1 49	0
controls	1 59	3 23	1 30	0
APM1				
subjects	3 26	2 53	3 4	5
controls	3 30	2 37	4 7	4
LEPA19G				
subjects	8 9	4 78	2 3	7
controls	7 9	4 98	3 8	3
MIF				
subjects	4 18	1 75	2 0	6
controls	3 94	1 98	2 2	4

Table 2. Summary of Random data results. In bold and in all capital letters, are the SNPs in question. The number below value 1 indicate the number of people who are homozygous for the wild-type (wt) and thus

do not express the SNP. The value 2 indicates number of people heterozygous. The value 3 indicates number of people homozygous for the variant (var) and thus do have the SNP and finally the value 4 indicates that the test failed.



Graph 1. Summarizes table 1. The type of SNPs and type of study subject are located on the X-axis. The subjects are those who have lymphoma, while the controls do not have lymphoma. The blue bar depicts the number of people homozygous for the wild-type (wt/wt) trait, red indicates number of people heterozygous (wt/var), yellow indicates number of people homozygous for the variant trait (var/var) meaning that they express the SNP.

	1	2	3	4
CY1PA				
subjects	0.24	0.52	0.24	0
controls	0.26	0.78	0.21	0
APM1				
subjects	0.53	0.41	0.05	0.009
controls	0.53	0.38	0.08	0.006
LEPA19G				
subjects	0.13	0.81	0.06	0.005
controls	0.13	0.77	0.04	0.01
MIF				
subjects	0.68	0.28	0.03	0.009
controls	0.64	0.32	0.03	0.006

Table 2. Summary of the ratios of people who are heterozygous or homozygous for the SNPs. The same values from Table 1 for the numbers 1 (wt/wt), 2 (wt/var), 3 (var/var), and 4 (fail) still apply.

SNP	TEST RESULTS
CY1PA	0.0015
APM1	0.0000675
LEPA19G	0.002
MIF	0.002
MIF & CY1PA	0.000266
MIF & APM1	0.002
MIF & LEP	0.00817

Table 4. A summary of the analysis using chi-squared, with 1 degree of freedom and a 95% confidence interval.

Discussion

Due to the difficulty of procuring the complete set of data, the results are very limited and very little can be concluded. The statistical analysis of the randomly generated data set implies that there is no relationship between developing lymphoma and presence of SNPs. From these results, the presence of one or more SNPs does not place a person in a high-risk category. According to these results, the initial goal of producing biomarkers is therefore not conceivable for this study. With results concluding a negative correlation between SNPs and lymphoma, we would want to consider future studies involving synergistic interactions between SNPs and risk factors, including cigarettes, alcohol intake and diet. Future studies could investigate high-risk lifestyles, whose high risk is due to the presence of the SNP. If a person is a factory worker and is exposed to high level of industrial chemicals, they might be placed in a high risk category, however, only if they have a certain SNP. Skibola *et al* accomplished studying lifestyle and lymphoma in the 1999 issue of *Proceedings of the National Academy of Sciences*, where they published results which provide evidence correlating a depletion in folic acid with lymphoma. The 1999 Skibola study also demonstrates that certain SNPs might actually help prevent against cancer in conjunction with a good diet.

In addition to the Skibola study in 1999, two studies conducted in 1993 and 1994 found that folic acid depletion is associated with an increased risk of colon cancer (Giovannucci *et al* 1993, Mason 1994). It has also been found that the Seventh Day Adventists, who are a religious group of non-smokers and vegetarians, have a longer life and half of the incidence of cancer than the average American (Fraser 1999), placing a high importance on not smoking and a diet rich in vegetables and fruits. With such

evidence it might be plausible that the incidence of lymphoma has more to do with diet or an interaction between diet and SNPs.

Future studies could also look at other factors, like race, sex, age, etc. In Duell's et al 2002 study of pancreatic adenocarcinoma, their results indicate that women are more susceptible to pancreatic cancer than men, due to differences in hormones. With women being more susceptible to pancreatic cancer, one might want to investigate any synergistic interactions between gender and the SNPs.

If the final results prove that there is no correlation between lymphoma and SNPs, it could be possible that diet or other factors are influencing the outcome. Therefore, it would be beneficial to investigate any synergistic interactions between people's diet, age and other risk factors and the presence of the SNPs

In addition to risk factors, there should also be consideration of other SNPs. If the SNPs in this study are found to have no correlation with lymphoma, this does not mean that all SNPs will have no correlation. Since there are about 1.8 million documented SNPs in the human genome (SNP consortium, 2003), future studies would want to focus their study on SNPs that have been found in other cancers or SNPs whose disruption in a specific biochemical pathway could lead to lymphoma. There are many that could be tested and probably many SNPs that are involved in the cancer development. Following Mimori *et al*'s 2002 study that found the SMARCB1 gene responsible for breast cancer, Comings *et al* found additional SNPs that were responsible for breast cancer. In addition to breast cancer, studies have looked at SNPs in pancreatic cancer, specifically SNPs within genes encoding for DNA repair enzymes (Duell *et al* 02). DNA repair enzymes are important with regards to correcting mistakes found in a DNA sequence. It is conceivable that cancer is caused by disruptions within the DNA repair enzymes. SNPs located within genes coding for the DNA repair enzymes are promising areas for future investigations with regards to lymphoma.

The above discussion focuses on the possibility that the results indicate no correlation between SNPs and the incidence of lymphoma. However, the other possibility is that there is indeed a correlation. With results indicating a positive correlation, the findings could assist in many areas of cancer treatment, including clinical diagnosis, classification and cancer drugs. The area of cancer drugs could be improved with the help of

developments in pharmogenomics due to discovery of SNPs. Pharmogenomics investigates how a person will respond to different drugs and improvements in this area would result in more personalized, safer and more efficient treatments (Saxena *et al* 03). Such optimized treatment regimes would allow an increase in the quality of life and expectancy for cancer patients.

In addition to improvements in cancer drugs, the SNPs would help in the classification and diagnosis of lymphoma possibly even placing certain people in a high-risk category. In other words, the SNPs could serve as biomarkers in the future.

Possible areas of error might include difficulty in investigating an equal distribution of ethnicities. There was an attempt to rectify this problem. However, to do this perfectly would be very difficult since there is a need for a very high volume of subjects. Lastly, of course there are sources of error due to human mistakes involved in the analysis, especially with such a high volume of people working on the study.

Acknowledgements

I would like to thank Martyn Smith's Molecular Epidemiology and Toxicology laboratory for letting me work in their lab. I would also like to thank my project supervisors John Latto, Matt Orr, and Donna Green. Thank you to the GSIs of ES 196: Manish Desai, Eric Dubinsky, Kevin Golden, and Renata Andrade

References

American Cancer Society (2002). http://www.cancer.org/downloads/STTcancer Facts&Figures2002TM.pdf

Cavalieri Ercole L., Li Kai-Ming, Bali Narayanan, Saeed Muhammad, Devanesan Prabu, Higginbotam Sheila, Zhao John, Gross Michail L., Rogan Eleanor G. Catechol ortho-quinones: the electrophilic compounds that form depurintating DNA adducts and could initiate cancer and other diseases. (02) Carcinogenisis Vol. 23 pp. 1071-1077.

Chang B, Heng SL, Isaacs SD, Turner A, Hawkins FA, Wiley KE, Bleecker ER, Walsh RC, Meyers DA, Isaacs WB, Xu J (03). Polymorphisms in the CYP1A Gene are associated with prostrate cancer risk. Int. J Cancer 106: 375-378.

Chesney J, Metz C, Bacher M, Peng T, Meinhardt A, Bucala R (98). An essential role for macrophage migration inhibitory factor (MIF) in angiogenesis and the growth of a murine lymphoma. Mol. Med. 5:179-189.

Comings De, Gade-Andavolu R, Cone LA, Muhlemand D, MAcMurray JP (03). A multigene test for risk of sporadic breast carcinoma. Cancer 97(9): 2160-2170

Diez JJ, Iglesias P (03). The role of the novel adipocyte-derived hormone adiponectin in human disease. European Journal of Endocrinology 148(3): 293-300.

Duell EJ, Holly EA, Bracci PM. Liu M, Wiencke JK, and Kelsey KT (02). A Population-based Study of Polymorphisms in Carcinogen-Metabolizing Genes, Smoking and Pancreatic Adenocarcinoma Risk. Journal of the National Cancer Institute 94: 297-306.

Duell EJ, Holly EA, Bracci PM. Wiencke JK, and Kelsey KT (02). A population-based Study of the Arg399Gln Polymorphism in X-ray Repair Cross-Complementing Group 1 (XRCC1) and Risk of Pancreatic Adenocarcinoma. Cancer Research 62: 4630-4636.

Edwards Y.H Potter, Hopkinson D.A. Biochem. Journal Vol. 187 pp. 429-436.

Eikelmann P,;Schulz W.A., Rohde D, Schmitz-Drager B., Sies H. Biol. Chem. Vol. 375 pp. 439-445.

Fraser GE (99). Association between diet and cancer, ischemic heart disease, and all cause mortality in non-Hispanic white California Seventh-day Adventists. American Journal of Clinical Nutrition 70(3): 532S-538S

History Etiology, Signs/ symptoms /risk factors of hodgkin's disease. http://pbl.cc.gatech.edu/mindy/847

Huggins, C. and Fukunishi, R. (64) Journal of Experimental Medicine 119: 923-942.

Kawakami D, Omura K, Kanehira E, Watanabe Y. (99) Polymorphic tandem repeats in the thymidylate sunthase gene is associated with its protein expression in human gastrointestinal cancers. Anticancer Research 19(4B): 3249-3252

Kim Y, Mason JB. Folate, epithelial dysplasia and colon cancer (95). Proc Assoc Am Physicians 107 (2): 218-227

Korte JE., Hertz-Picciotto I, Schulz MR ,Ball LM, Duell EJ (00). The Contribution of Benzene to Smoking-Induced Leukemia. Environmental Health Perspectives 108: 333-339.

Kutyavin IV, Afonina IA, Mills A, *et al*. (00). 3' minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. Nucleic Acids research 28: 655-661.

Mimori K, Inoue H, Shiraishi T, Ueo H, Mafune K, Tanaka Y and Mori M (02). A Single-Nucleotide Polymorphism of SMARCB1 in Human Breast Cancers. Genomics 80: 254-258.

Mitchel RA, Bucala R. Tumor growth – promoting properties of macrophage migration inhibitory factor (MIF)

Neber DW, Roe AL, Vandale SE., Bingham E, Oakley GG (02). NAD (P) H: Quinone Oxidoreductase (NQO1) Polymorphism, Exposure to Benzene, and Predisposition to Disease. Genetics in Medicine Vol. 4 pp. 62-70.

Real Time PCR Goes Prime. TechNotes 8(1). http://www.ambion.com/techlib

Rothman N, Smith Martyn T, Hayes RB, Li GL, Irons RD, Dosemeci M, Haas R, Stillman WS, Linet M, Xi LQ, *et al*. (96). Environmental Health Perspectives Vol. 104 pp. 1365-1370.

Saxena A, Agarwal S. and Rao YN (03). Molecular Approaches to Cancer Management in the Post-genomic Era. http://mohfw.nic.in/pg44to48.pdf

Skibola CF, Smith MT, Kane E, Roman E, Rollison S, Cartwright RA, Morgan G (99). Polymorphisms in the methylenetrahydrofolate reductase gene are associated with susceptibility to acute leukemia in adults. Proc. Natl. Acad. Sci. 96: 12810-12815.

Skibola CF, Smith MT, Hubbard A, Shane B, Roberts AC, Law GR, Rollinson S, Roman E, Cartwright RA, Morgan GJ (02). Polymorphisms in the thymidylate synthase and serine hydroxymethyltransferase genes and risk of adult acute lymphocytic leukemia. Neoplasia. 99: 3786-3791

Smith MT. Benzene, NQO1, and genetic susceptibility to cancer. (99) Proc. Natl. Acad. Sci. 96:. 7624-7626.

Smith MT, Wang Y, Kane E, Rollinson S, Wiemels JL, Roman E, Roddam, Cartwright R, Morgan G (01). Low NAD(P)H: quinone oxidoreductase 1 activity is associated with increased risk of acute leukemia in adults. Blood 97: 1422-1426.

Smith MT. The Mechanism of Benzene-induced Leukemia: A Hypothesis and Speculations on the Causes of Leukemia. (96) Environmental Health Perspectives 104: 1219-1225.

SNP consortium: website http:snp.cshl.org/

Traver R.D., Horikoshi T, Danenberg K.D., Stadlbauer T. H., Danenberg P.V., Ross D., Gibson N.W. (92) Cancer Research Vo. 52 pp. 797- 802.

Zhang L, Rothman N, Wang Y, Hayes RB, Bechtold W, Venkatesh P, Yin S, Wang Y, Dosemeci M, Li G, Lu W, Smith MT (96). Interphase Cytogenetics of Workers Exposed to Benzene. Environmental Health Perspectives 104: 1325-1329.

Zheeb H, Blettner M (98). Adutl leukemia: what is the role of currently known risk factors? Radiat Environ Biophys. 36: 217-228