Environmental Etiology of Polycythemia Vera, Essential Thrombocythemia and Primary Myelofibrosis: A Case-Control Study in Northeast Pennsylvania

A Thesis

Submitted to Faculty

Of

Drexel University

by

Carol Ann Gross-Davis

in partial fulfilment of the

requirements for the degree

of

Doctor of Philosophy

September 2013

Dedication

For Sophia and Henry,

Thank you for giving me the time, and the encouragement to never give up.

Acknowledgments

I would like to offer my gratitude to a special few individuals for their support throughout this journey. My Department Chair, Dr. Arthur L. Frank, has been a mentor and has provided to me the resources and expertise for this study to be possible. My advisors, Dr. Craig Newschaffer and Dr. Igor Burstyn, both exceptional teachers, who pushed me and made this experience a thoughtful journey. Their patient guidance and constructive criticisms at different stages of my research was truly invaluable. Thank you Dr. Igor Burstyn for the friendly and thought provoking questions and encouragement, I have learned so much from our conversations. Dr. Karyn Heavner for her daily support in data cleaning and analysis and of course her friendship. Professor Richard Pepino for his continued guidance, mentorship and friendship he has provided to me since my undergraduate studies. I would also like to thank members of my dissertation committee -Drs. Allison Evans and Palak Ravel-Nelson, for their help over the years as my ideas evolved to into a completed study. And my friends, especially Dr. Nicole Gidaya for her kindness and support and in always finding some fun in this pursuit. I would also like to thank my family –Mom, Dad, and Hugh for always pushing me to do my best, and letting me reaching for the stars. I also want to thank Sophia and Henry, for constantly encouraging me every step of the way. I am deeply grateful to have you all, especially my father, who anchored the support system that allowed me to stay focused on my graduate study. I truly could not have completed this without you all, and your belief in me.

This final report (*Investigation of a Polycythemia Vera, Essential Thrombocythemia and Primary Myelofibrosis Cluster in Northeast Pennsylvania: A Case-Control Study*) was supported by a Grant/Cooperative Agreement/Earmark from the Center for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of ATSDR.

Table of Contents

List of Tables

Abstract

Environmental Etiology of Polycythemia Vera, Essential Thrombocythemia and Primary Myelofibrosis: A Case-Control Study in Northeast Pennsylvania

Carol Ann Gross-Davis, MS

Advisors: Craig Newschaffer, PhD., Igor Burstyn PhD

Objectives

The etiology of a rare category of myeloproliferative neoplasms (MPN), bone marrow diseases with an excess of blood cells, is currently unknown. An MPN cluster in Northeastern Pennsylvania allowed investigation of effects of environmental risk factors and to assess the potential for gene-environment interactions. Since no strong hypothesis could be advanced, we focused on known occupational and environmental carcinogens, especially those previously implicated in blood tumors, in our investigation (e.g. benzene, polycyclic aromatic hydrocarbons (PAH), aromatic & heterocyclic amines). The aims of this dissertation were to evaluate the associations between lifestyle and environmental risk factors for the most common myeloproliferative neoplasms (MPNs) (PV, ET, and PMF), with and without JAK2 V617FV617F. We also explore an interaction between known susceptibility genotypes for a subset of cases and controls and potential mutagenic chemical exposures, including PAHs.

Methods

This 2011 population based case-control study assessed residential, smoking and dietary history by telephone interview in Schuylkill, Luzerne and Carbon counties in Pennsylvania. Cases (n=55) were identified from the Pennsylvania cancer registry and a previous MPN study in Pennsylvania. Controls (n=473) were selected based on eligibility screening using random digit dialing. Blood samples for genotyping were collected from a subset of 31 cases and 292 controls.

Results

Cases were older (median age=71 vs. 61years) and more likely to be male (49% vs. 39%) compared to controls but otherwise demographically similar. We found no relationships between MPNs and smoking history, residential history, diet, and lifestyle behaviors with presumed exposure to aromatic and heterocyclic amines. After studying the main effects of 14 environmentally sensitive genes, we found that only the NAT2, CYP1A2, GSTA1, and GSTM3 variants were associated with an average of 3-to 5-fold increased odds of having an MPN.

Conclusion

While these results do not confirm a gene environment interaction effect for one specific chemical, the findings encourage further exploration of the interaction hypothesis with respect to NAT2, GST, and CYP gene biological pathway and chemical exposures. These same genes appear to be implicated in cases with the somatic mutation believed to be involved in the etiology of MPNs, especially PV. Although no association was found between exposures related to aromatic and heterocyclic amines and MPNs, our findings suggest that genotypes that modify the toxicity of these exposures may play a role in MPNs.

Chapter 1: Background and the Overview of the Thesis

Myeloproliferative neoplasms (MPNs)

MPNs are diseases of the bone marrow, with *myeloproliferative* meaning the excess of blood cells. MPNs are characterized by an overproduction of mature erythrocytes, often platelets (James et al. 2005). These diseases were previously called *myeloproliferative disorders* and the classification of their related diseases has changed numerous times.

MPNs as classified by the World Health Organization (WHO) include polycythemia vera (PV), essential thrombocythemia (ET), and primary myelofibrosis (PMF). PV is a condition that results in the overproduction of red blood cells, >25% above predicted red cell mass (Campbell & Green, 2005). ET is an overproduction of platelets in the bone marrow. PMF includes bone marrow fibrosis (or scarring of the bone)and is the most severe of the three (Tefferi & Vardiman, 2007) .

In 2008, the WHO's revised diagnostic criterion to include molecular testing for MPN diagnosis. Prior to the revised diagnostic criteria in 2001 Kutti and Ridell calculated a yearly incidence of PV at 2.8 per 100,000 persons per year (Kutti &Ridell, 2001). For ET the incidence rate was estimated at 1.5 cases per 100,000 per year (Kutti &Ridell, 2001). The rarest of the MPNs is PMF, with an incidence rate of only 0.4 cases per 100,000 persons per year (Kutti &Ridell, 2001).

Incidence and prevalence reporting of these MPNs are improving each year. However, after only one decade of reporting and diagnostic criteria evolution, there is still an ongoing effort by the CDC to improve reporting. Since 2001, PV, ET, and PMF were required to be reported to the state cancer registries (Seaman et al., 2009). However,

only hospital cases were required to be reported and, because treatments of MPNs typically do not require hospitalization, physicians did not report cases to the Cancer Registries such as Pennsylvania Cancer Registry (PCR). In 2006, Seaman et al. found 53% (20 of 38) of the cases that participated and were reported to the PCR either did not have enough information to confirm a PV diagnosis or did not meet the case definition of PV.

The outcomes and demographics for patients with PV, ET, and PMF differ. Patients with PV and ET can expect to have a near-normal life expectancy if the disease and potential complications are managed properly. PV patients may progress to myelofibrosis and acute leukemia (Tefferi, 2007). For PV the median age of diagnosis is 60 years, with a slight male predominance (Tefferi, 2007).. However, the outcome for PMF patients is worse than for PV and ET. PMF patients have a lower quality of life and survival rate. No effective therapy is available for PMF, with survival of 60 % at five years (Tefferi & Vainchenker, 2011).

Janus Kinase 2 (JAK2 6V17F) point mutation

An acquired somatic point mutation in Janus Kinase 2 (JAK2 6V17F) on chromosome 9p, is seen in nearly all PV patients (>95%) and about half of those with ET or PMF. This mutation is suspected to be pathognomonic. JAK2 is a signalling protein that acts as an on-and-off switch regulating bone marrow activity. This mutation disrupts the normal inhibition of growth, thus increasing blood cell production (Baxter et al., 2005). While the MPN types differ phenotypically, they commonly share the same de-novo somatic JAK2 V617F point mutation (JAK2) thought to be at least partly responsible for disease initiation and/or progression (Baxter et al., 2005).

The etiologies of MPNs and the causes of JAK2 V617Facquired mutations are unclear (Baxter, 2005). The WHO included the JAK2 V617F mutation as a major diagnostic criterion for PV, ET, and PMF in 2008, although confirmed cases of PV, ET, and PMF can be JAK2 V617Fnegative if other criteria are satisfied (Seaman et al., 2009; Tefferi et al., 2007).

Familial clustering of PV, ET, and PMF

There is strong evidence for an increased risk of developing an MPN in people who have a family member with an MPN (Anderson et al., 2012, Landgren et al., 2008). Familial clustering has been documented in a large Swedish population-based study of 24,577 first-degree relatives of 11,039 MPN patients diagnosed from 1958 to 2005. This study showed a 5-7-fold increase in risk of developing an MPN for first-degree relatives of MPN patients (Landgren et al., 2008). However, this study could include false positive cases, as it included cases before the molecular diagnostic tool for the JAK2 V617F mutation was used.

Inherited genetics have also been suspected to influence both MPN phenotype and susceptibility, including potential germ-line mutations, some yet to be identified (Andrikovics et al., 2010). Recently, a germ-line haplotype, a sequence of single nucleotide polymorphisms (SNPs) on the loci designated as the 46/1 haplotype, has been identified as strongly associated with JAK2 V617Fstatus in chromosome 9p (Jones et al., 2009; Kilpivaara et al., 2009) (Olcaydu et al., 2009). Three genome-wide studies have reported an association with this germ-line 46/1 haplotype in European population and JAK2 V617F MPNs (Jones et al., 2009; Kilpivaara et al., 2009; Olcaydu et al., 2009). The 46/1 haplotype is estimated to be present in approximately 50% of the population and, for those with the JAK2 V617F mutation, is associated with a three-fold increase in

risk of developing an JAK2 V617Fassociated MPN *(*Jones et al.,2009*)*. The 46/1 haplotype is considered to be the only known risk factor for the JAK2 V617F mutation in PV patients (Andrikovics et al., 2010; Jones et al., 2009).

MPNs and occupational and chemical exposures

Starting with the discovery of the JAK2 V617F mutation and the 46/1 haplotype mutation, interest in environmental exposures and MPNs has emerged (Anderson et al.,2012). There is evidence in the literature to suggest that some occupational chemical exposures may be associated with MPNs.

Occupational exposures suggest petroleum products or benzene as a risk factor

Occupational groups have been investigated for specific chemical exposures, but there is no consensus on what specific occupational exposures are associated with MPNs. Exposure to petroleum, of which benzene is a by-product, was associated with an increased risk of PV and myelofibrosis (demonstrated by Kaplan (1986)), although there were only seven cases. Kaplan reported on a large cohort of petroleum refinery workers and found an elevated SMR in petroleum refiner workers (455, 95% CI: 120, 1164) (Kaplan, 1986). The SMR for myelofibrosis was also elevated for these workers at 201 (95% CI: 41, 588). This data only included four deaths from PV and three deaths from myelofibrosis making SMR estimates imprecise. Further investigation of PV or myelofibrosis was not recommended due to the small numbers of cases (Kaplan, 1986). All three of the above cohort studies suffered from a small numbers of cases. Only six were PV out of a total of 14 MPN reported cases, and there were no ET or PMF cases.

Two additional case-control studies done by Terreros et al. (1997) and Mele et al. (1997) on environmental and occupation exposures also suggest petroleum products or benzene

as a risk factor. Only Terreros, with nine cases of myeloproliferative syndrome, reported a statistically significant association (OR =47, 95% CI: 2, 2761) with self-reported exposure to benzene.

A study in 1981 in Argentina reported an elevated risk of hemopoietic system neoplasms with exposure to petroleum (Quiroga et al., 1981). Quiroga et al. (1981) conducted a case-control study from 16 medical centers in Argentina on 603 age matched cases and controls. In this hospital-based study, cases were diagnosed with hemopoietic system neoplasms (HSN) between 1973 and 1980. The analysis reported a relative risk (RR) of 7.84 with no confidence intervals reported for myeloproliferative disorders with exposure to petroleum products. The total MPNs were 51, although only ten PV cases and four myeloproliferative disorder cases were included (Quiroga et al., 1981). This study was conducted a few decades ago without the advanced ability to detect a cancer and the pathways for their causes (let alone a bone marrow malignancy) does call into question the usefulness of such studies. For instance, Quiroga et al. also focused on factors such as eye color and co-habitation with dogs (RR=1.30 for all HSNs).

Risk of hematological malignancies and exposure to PAH was also observed in a case control study in Italy but not specifically for myeloproliferative syndromes. Pasqualetti et al. (1991) reported an association for risk of hematological malignancies and exposure to aromatic hydrocarbons (OR=2.5, 95%CI: 1.39, 3.32). This study consisted of 620 cases of hematological malignancies and 1,240 age- and sex-matched controls. The controls were from the same geographical area as the cases with other medical conditions, except for cancer or congenital disease. Exposures were assessed by self-report during an interview. No association was found with myeloproliferative syndrome cases (n=44, ORs not reported) (Pasqualetti et al., 1991).

Schnatter et al. (1996) updated three nested case–control studies in 2012 specifically looking at occupational benzene exposure and lympho-hematopoetic (LH) (Schnatter, 1996, Rushton, 1997, Glass, et al., 2003). The three cohort studies that were used for the nested case-control study evaluated mortality in three different exposed populations. The first was mortality in Canadian petroleum workers from January 1964 to December 1983. The second was on Australian petroleum workers, looking at mortality and cancer incidence from January 1981 through December 1996. The last cohort looked at mortality and cancer incidence in UK petroleum workers from January 1950 to December 1989. In the 2012 Schnatter report, these nested case-control studies were updated and new cases of LH were added using mortality and cancer incidence registries from the respected countries of origin. Each study included at least ten additional years of follow-up for cases (Canada: December 1994; Australia: December, 2006; and the UK: December, 2005). The resulting data was pooled to look at the risk of five specific lymphatic–hematopoietic subtypes and benzene exposure. The LH subtypes included were three types of leukemia-acute myeloid leukemia (AML), chronic myeloid leukemia (CML), and chronic lymphocytic leukemia (CLL) and two types of myeloid neoplasm, myelodysplastic syndrome (MDS) and myeloproliferative disorders (MPD) (now referred to as MPNs). This pooled study population found an OR= 1.79 (95% CI: 0.68, 4.74) for 30 cases of MPD with cumulative exposure assessed as the highest exposure to benzene (>=2.93 vs. <= 0.348 ppm-years) (Schnatter et al., 2012).

All of these studies are of limited size and none of the studies mentioned above used the JAK2 V617F mutation molecular diagnostic tool, which is standard practice for diagnosis in 2013 (Seaman et al., 2009). The lack of a molecular diagnostic tool affects their ability to assess error due to case misclassification, which has been documented to be upwards of 50% (Seaman et al., 2010). As with many occupational studies, the lack of

information about lifestyle exposures is often not complete or non-existent in some cases, as in the Quiroga study of 1981. Even the most recent of studies, the Schnatter study in 2012 investigating benzene exposures, did not use the latest WHO 2008 diagnostic criteria. Schnatter only included lifestyle exposures such as smoking history for half of the studies cases, and no information on family history or diet was included.

Two cohort studies looking at PV and/or myelofibrosis reported higher mortality rates among different occupational groups involving exposures other than petroleum or benzene exposure (Zoloth et al., 1986, Johnson et al., 2010, Kaplan, 1986). In 1986, Zoloth reported higher than expected standardized mortality rates (SMRs) among 2,500 commercial pressmen in three study locations in the US: New York City, Long Island, and New Jersey. This cohort used incident cases from 1958 to 1981, with follow up for mortality until 1981. They reported an annual incidence rate of 120 cases per 100,000, compared to a population incidence in the population of 2 per 100,000 with only three cases of reported of myelofibrosis in the study population (Zoloth, 1986).

In a study of 2,639 local union members whom ever worked in poultry processing plant, Johnson et al., reported higher PV along with excess mortality of lymphoid leukemia and myelofibrosis mortality rates for a cohort of poultry workers. A proportional mortality rate of 4.9 (95% CI: 1.4, 17.2) and 10.9 (95% CI: 1.4, 85.0) for non-white females (Johnson et al., 2010) was also reported. However with only two cases of myelofibrosis, the precision of this study was impacted and should be noted. Johnson et al. (2010) then updated their previous mortality study of poultry workers to consider exposure to oncogenic viruses.

Environmental exposures and blood cancer

A large number of studies indicate exposure to many different chemicals may be associated with an increased risk of leukemia. Leukemia, like MPNs, is a disease of the bone marrow. However, the literature is inconsistent on providing evidence that exposure to individual agents (for instance, pesticides, metals, dioxins, and even PAHs) is associated with an increased risk of leukemia (Clapp et al., 2008). In the Agricultural Health Study, risk doubled among a cohort of pesticide applicators with exposure to organochlorine pesticides like chlordane, aldrin, and toxaphene (Clapp et al., 2008; Purdue et al., 2007). Studies looking at reactive chemicals are limited, but in one occupational study of synthetic rubber workers, strong associations between both high levels and intensity of butadiene exposure with leukemia were found (Alder et al., 2006). However, due to the many chemicals these workers are exposed to, it was not possible to relate the increased death meta-SMR of 1.2 specifically to exposure to butadiene (Alder et al., 2006; Clapp et al., 2008). The potential role for benzene and its causal relationship to acute leukemia has been reviewed elsewhere by Smith, MT (2011), and Smith MT and Zang, (2011). They indicate that benzene is associated with myelodysplastic syndromes (MDS) and acute myeloid leukemia (AML). This supports the biological plausibility that mutagenic chemicals like benzene are toxic to hematopoietic stem cells (HSC) or progenitor cells, from which these bone-marrow disorders arise (Smith, 2010; Smith et al., 2011). Acute and chronic exposures and observed hematologic effects have been well studied to show their direct association with hematologic changes in humans as well as animals (Galbraith et al., 2010; Smith, 2010).

Investigating mutagenic exposures and their ability to make changes in our DNA is part of a growing body of literature. Still much more research is needed, in particular, large

studies estimating the magnitude of effects estimates with more precision, as well as studies obtaining more accurate concentrations of exposures.

PAHs and risk factors for leukemia

Recent studies of exposure to PAHs and risk factors for childhood leukemia provide support that changes in bone marrow resulting from a chemical exposure may have a role in chronic leukemia diseases like MPNs (Rull et al., 2008). In a recent sub-study of the Northern California Childhood Leukemia Study, PAH exposure was measured in the residential indoor dust in homes of 227 controls and 210 cases with childhood leukemia (Rull et al., 2008). In single-PAH risk models comparing highest to lowest quartiles, they observed elevated odds ratios for three PAHs: benzo[k]fluoranthene (1.8, 95% CI: 1.0-3.5); benzo[b]fluoranthene (2.0, 95% CI: 1.0-3.9); indeno[1,2,3-cd]pyrene (1.7, 95% CI: 0.9-3.2) (Rull et al., 2008).

A case-control study of acute myeloid leukemia (AML) in an eastern Indian population included 110 AML cases and 144 geographically and racially matched healthy controls. It showed an increase in risk for developing AML with an odds ratio of 11.91 (CI: 4.04-34.96, $p<0.001$) in subjects with both Glutathione S-transferase (GST) polymorphism GSTM1"null" and Polymorphic N-acetyltransferase (NAT2) (Majumdar et al., 2008). These findings suggest that benzo[a]pyrene exposure is associated with the elevated risk seen in both GSTM1 null and NAT2 genotypes.

Although many studies group these hematological malignancies together with MPNs, they are not the same disease. Specifically, one major difference is that in the classic MPNs (PV, ET, and PMF), patients do not carry the Philadelphia chromosome—they are (PH) negative. The Philadelphia chromosome adversely impacts normal processes in the bone marrow cells and produces an abnormal protein, a different mechanism than found in MPNs. This results in the proliferation of leukemic cells. The Philadelphia

chromosome is also present in some cases of CML, AML, and ALL, which contributes to a more unfavorable prognosis with increasing age.

Genetic susceptibility and PV, ET, PMF, and the JAK2 V617F mutation

One family of genes that has been long studied and linked to increased cancer risk in general are the CYPs, GST, and NAT2 genes. Genetic susceptibility might explain heterogeneity of environmental exposure effects on cancer risk. It might also explain why some individuals have a disease risk at exposure levels that would not cause adverse effects for the majority of people. However, there may also be other important categories of cancer susceptibility genes.

The National Institute of Environmental Health Sciences Environmental Genome Project (NIEHS EGP) was developed in 1997 to explore and categorize environmentally responsive genes (ERG) in the human genome. The NIEHS EGP finished gene list included 648 genes in the following eight categories: cell cycle; cell division; cell signalling; cell structure; DNA repair; gene expression; homeostasis and metabolism. Although these genes are being studied for a wide range of cancers, no studies have investigated any combination of these genotypes and the risk of MPNs. Because JAK2 V617Fis a *de novo* mutation of JAK2 V617Fand its role in MPNs, the gene-environment interaction mechanisms of most interest here relates to genes that modify susceptibility to mutagenic chemicals.

Conclusion

Environmental exposure and MPNs are not well enough studied to help us understand how these exposures may play a role in the etiology of the disease (Anderson et al., 2012). As described previously, there is evidence that chemicals can penetrate the cells

in the bone marrow and this is consistent with the biological pathways considers in this study (Smith, 2010; Smith et al., 2010). In this investigation we focused on known mutagenic chemicals similar to benzene and PAHs, and the susceptible genotypes that modify these exposures, especially those previously implicated in blood cancer.

Aims and Overview of the Dissertation

Introduction

In 2008, the Pennsylvania Department of Health (PADOH) and the Agency for Toxic Substances and Registry (ATSDR) of the Centers for Disease Control and Prevention (CDC) confirmed a PV cluster in a tri-county region of Northeast Pennsylvania. The incidence of PV was 4.3 times higher within this cluster area than in a comparative tri-county region. The close proximity of this primary cluster zone—located at the intersection of Carbon, Luzerne, and Schuylkill Counties—to known hazardous waste sites and co-generation power plants raised the question as to whether they could have played a causal role in the disease cluster (Seaman et al., 2009).

The aims of this dissertation are to evaluate the associations between lifestyle and environmental risk factors for the most common MPNs (PV, ET, and PMF), with and without JAK2 V617F. We also explore an interaction between known susceptibility genotypes for a subset of cases and controls and potential mutagenic chemical exposures, including PAHs.

Aims

Aim 1: Estimate the effect of exposure to PAHs on risk of MPNs in the tri-county area of Northeast Pennsylvania.

Aim 1a: Estimate the effect of exposure to PAHs from residential sources, diet, and lifestyle/behavioral factors on the risk of PV-related outcomes.

Aim 2: Explore the role of genotypes that may modify the potential of mutagenic chemicals to affect the risk for MPNsin the tri-county area.

Aim 3: Estimate the joint effects of susceptible genotypes identified in Aim 2 and distance of residence from facilities with known hazardous emissions on risk of MPNs.

The Drexel University School of Public Health, Department of Environmental and Occupational Health received funding from the CDC to investigate risk factors associated with PV in the tri-county area. This Drexel University study was meant to be a population-based, case-referent study to evaluate possible lifestyle, environmental, and genetic risk factors for a family of myeloproliferative neoplasms (MPNs) believed to be related to PV, which include essential thrombocythemia (ET) and primary myelofibrosis (PMF).

Organization of this dissertation

There are five chapters in this dissertation. The overview and review of relevant literature of the dissertation was included in this chapter. Chapter 2 addresses Aim 1 by describing a population-based, case-referent study that evaluated lifestyle and environmental risk factors for PV, ET, and PMF, as well as the risk factors for JAK2 V617F. In chapter 3, we use a subset of that study sample to investigate presence (in a qualitative sense) of gene-environment interactions, using gene-only analysis (Burstyn et al., 2009) for Aim 2. We also investigate the association between the susceptible genotypes that can disrupt the metabolic breakdown of mutagenic chemicals with the risk of an MPN. The genes we investigated were AHR, CYP1A1, CYP1A2, CYP1B1, CYP2B6, GSTP1, GSTM1, GSTM3, GSTT1, NAT2, and NQO1. We hypothesized in Aims 1 and 2 that there is increased risk of an MPN associated with aromatic and heterocyclic amines, as well as benzene. In Chapter 4, we explore the possibility of gene-environment interaction by examining the identified genotypes in Aim 2 and chemical exposures assessed by using distance to hazardous facilities as a proxy for exposure. The dissertation findings are summarized in chapter 5.

This dissertation was designed to explore the environmental etiology of polycythemia vera, essential thrombocythemia, and primary myelofibrosis. Results from this dissertation add to the very limited amount of epidemiological studies looking at environmental exposures as risk factors to MPNs. It will encourage further research for chemical exposures as well as susceptible genotypes for these diseases.

List of References

Anderson, L. A., Duncombe, A. S., Hughes, M., Mills, M. E., Wilson, J. C., & McMullin, M. F. (2012). Environmental, lifestyle, and familial/ethnic factors associated with myeloproliferative neoplasms. *Am J Hematol, 87*(2), 175-182.

Alder, N., Fenty, J., Warren, F., Sutton, A. J., Rushton, L., Jones, D. R., & Abrams, K. R. (2006). Meta-analysis of mortality and cancer incidence among workers in the synthetic rubber-producing industry. *Am J Epidemiol, 164*(5), 405-420.

Andrikovics, H., Nahajevszky, S., Koszarska, M., Meggyesi, N., Bors, A., Halm, G., Tordai, A. (2010). JAK2 46/1 haplotype analysis in myeloproliferative neoplasms and acute myeloid leukemia. *Leukemia, 24*(10), 1809-1813.

Burstyn, I., Kim, H.-M., Yasui, Y., & Cherry, N. M. (2009). The virtues of a deliberately mis-specified disease model in demonstrating a gene-environment interaction. Occupational and Environmental Medicine, 66(6), 374-380.

Campbell, P. J., & Green, A. R. (2005). Management of polycythemia vera and essential thrombocythemia. *Hematology Am Soc Hematol Educ Program*, 201-208.

Clapp, R. W., Jacobs, M. M., & Loechler, E. L. (2008). Environmental and occupational causes of cancer: new evidence 2005-2007. *Rev Environ Health, 23*(1), 1-37.

Falcetta R,Sacerdote C,Bazzan M, et al. Occupational and lifestyle risk factors for essential thrombocytemia: A case–control study. G Ital Med Lav Ergon 2003; 25: 9–12.

Finazzi, G., Rambaldi, A., Guerini, V., Carobbo, A., & Barbui, T. (2007). Risk of thrombosis in patients with essential thrombocythemia and polycythemia vera according to JAK2 V617F mutation status. *Haematologica, 92*(1), 135-136.

Galbraith, D., Gross, S. A., & Paustenbach, D. (2010). Benzene and human health: A historical review and appraisal of associations with various diseases. *Crit Rev Toxicol, 40 Suppl 2*, 1-46.

Glass DC, Gray CN, Jolley DJ, et al. Leukemia risk associated with low-level benzene exposure. Epidemiology 2003. 14(5):569–577.

Johnson ES, Zhou Y, Lillian Yau C, et al. Mortality from malignant diseases- update of the Baltimore union poultry cohort. Cancer Causes Control 2010;21:215–221.

Jones, Amy V., Chase, Andrew, Silver, Richard T., Oscier, David, Zoi, Katerina, Wang, Y. Lynn,. Cross, Nicholas C. P. (2009). JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nat Genet, 41*(4), 446-449.

Kilpivaara, Outi, Mukherjee, Semanti, Schram, Alison M., Wadleigh, Martha, Mullally, Ann, Ebert, Benjamin L.,Levine, Ross L. (2009). A germline JAK2 SNP is associated with predisposition to the development of JAK2V617F-positive myeloproliferative neoplasms. *Nat Genet, 41*(4), 455-459.

Landgren, O., Kristinsson, S. Y., Goldin, L. R., Caporaso, N. E., Blimark, C., Mellqvist, U. H., Turesson, I. (2009). Risk of plasma cell and lymphoproliferative disorders among 14621 first-degree relatives of 4458 patients with monoclonal gammopathy of undetermined significance in Sweden. *Blood, 114*(4), 791-795

Majumdar, S., Mondal, B. C., Ghosh, M., Dey, S., Mukhopadhyay, A., Chandra, S., et al. (2008). Association of cytochrome P450, glutathione S-transferase and N-acetyl transferase 2 gene polymorphisms with incidence of acute myeloid leukemia. *Eur J Cancer Prev, 17*(2), 125-132.

Mele A, Visani G, Pulsoni A, et al. Risk factors for essential thrombocythe- mia—A case–control study. Cancer 1996;77:2157–2161.

Mesa, R. A., Verstovsek, S., Cervantes, F., Barosi, G., Reilly, J. T., Dupriez, B., Tefferi, A. (2007). Primary myelofibrosis (PMF), post polycythemia vera myelofibrosis (post-PV MF), post essential thrombocythemia myelofibrosis (post-ET MF), blast phase PMF (PMF-BP): Consensus on terminology by the international working group for myelofibrosis research and treatment (IWG-MRT). *Leuk Res, 31*(6), 737-740.

Moliterno, A. R., Williams, D. M., Rogers, O., Isaacs, M. A., & Spivak, J. L. (2008). Phenotypic variability within the JAK2 V617F-positive MPD: roles of progenitor cell and neutrophil allele burdens. *Exp Hematol, 36*(11), 1480-1486.

Olcaydu, Damla, Harutyunyan, Ashot, Jager, Roland, Berg, Tiina, Gisslinger, Bettina, Pabinger, Ingrid, Kralovics, Robert. (2009). A common JAK2 haplotype confers susceptibility to myeloproliferative neoplasms. *Nat Genet, 41*(4), 450-454.

Passamonti, Francesco, & Rumi, Elisa. (2009). Clinical relevance of JAK2 (V617F) mutant allele burden. *Haematologica, 94*(1), 7-10.

Pasqualetti P, Casale R, Colantonio D, Collacciani A. Occupational risk for hematological malignancies. Am J Hematol 1991;38:147–149.

Purdue, M. P., Hoppin, J. A., Blair, A., Dosemeci, M., & Alavanja, M. C. (2007). Occupational exposure to organochlorine insecticides and cancer incidence in the Agricultural Health Study. *Int J Cancer, 120*(3), 642-649.

Rull, R. P., Gunier, R. B., Reynolds, P., Colt, J. S., Nishioka, M., Metayer, C., et al. (2008). Residential Concentrations of Polycyclic Aromatic Hydrocarbons and Childhood Leukemia Risk. *Epidemiology, 19*(6), S275

Rushton L, Romaniuk H. A case-control study to investigate the risk of leukaemia associated with exposure to benzene in petroleum marketing and distribution workers in the United Kingdom. Occup Environ Med 1997. 54(3):152–166.

Schnatter AR, Armstrong TW, Nicolich MJ, et al. Lymphohaematopoietic malignancies and quantitative estimates of exposure to benzene in Canadian petroleum distribution workers. Occup Environ Med 1996. 53(11):773–781.

Schnatter, Robert Glass, Deborah C. Tang, Gong, Irons, Richard D. and Rushton,Lesley. (2012) Myelodysplastic (2012). Syndrome and Benzene Exposure Among Petroleum Workers: An International Pooled Analysis. JNCI J Natl Cancer Inst.

Seaman, V., Jumaan, A., Yanni, E., Lewis, B., Neyer, J., Roda, P., Hoffman, R. (2009). Use of molecular testing to identify a cluster of patients with polycythemia vera in eastern Pennsylvania. *Cancer Epidemiol Biomarkers Prev, 18*(2), 534-540.

Smith, Martyn T. (2010). Advances in Understanding Benzene Health Effects and Susceptibility. *Annual Review of Public Health, 31*(1), 133-148.

Smith, Martyn T., Zhang, Luoping, McHale, Cliona M., Skibola, Christine F., & Rappaport, Stephen M. (2011). Benzene, the exposome and future investigations of leukemia etiology. *Chemico-Biological Interactions, In Press, Uncorrected Proof*.

Spivak JL. (2010) Narrative review: Thrombocytosis, polycythemia vera, and JAK2 mutations: The phenotypic mimicry of chronic myelofibrosis. Ann Intern Med. 2010 Mar 2;152(5):300-6. Review.

Tefferi, A., & Vainchenker, W. (2011). Myeloproliferative neoplasms: molecular

pathophysiology, essential clinical understanding, and treatment strategies. *J Clin Oncol, 29*(5), 573-582.

Tefferi, A., & Vardiman, J. W. (2007). Classification and diagnosis of myeloproliferative neoplasms: The 2008 World Health Organization criteria and point-of-care diagnostic algorithms. *Leukemia, 22*(1), 14-22.

Terreros MC, Apezteguia M, Slavutsky IR, Guimarey LM.(1997) Exposure to occupational and environmental risk factors in hematologic disorders. Neoplasia;14:133–136.

Quiroga Micheo E, Calcagno EJ, Calabria SI, et al. (1981). Retrospective epidemio- logical study of hemopoietic system neoplasms in Argentina. Medicina;41:187–200.

Zoloth SR, Michaels DM, Villalbi JR, Lacher M. (1986).
Patterns of mortality among commercial pressmen. J Natl Cancer Inst. 1986 Jun;76(6):1047-51.

Chapter 2: The Risk of Myeloproliferative Neoplasms Due to Exposure to Polycyclic Aromatic Hydrocarbon

Abstract

Background

The etiology of myeloproliferative neoplasms (MPN), rare bone marrow diseases with an excess of blood cells, is currently unknown. A somatic point mutation—JAK2 V617F—is suspected to be an antecedent of the most common MPN, polycythemia vera (PV). This mutation may also occur in excess in related MPNs (essential thrombocythemia (ET) and primary myelofibrosis (PMF)). Consequently, environmental exposure to mutagens, such as polycyclic aromatic hydrocarbons (PAH) and heterocyclic amines, is suspected to contribute to PV. An MPN cluster in northeastern Pennsylvania allowed investigation of residential and lifestyle risk factors, with particular focus on exposure to PAH.

Objectives

We evaluated exposure to PAH as a risk factor for PV, ET, and PMF. We also studied the risk factors for having the JAK2 V617F mutation.

Methods

We conducted a population-based case-control study in northeastern Pennsylvania. People born from 1921-1968 and residing between 2000 and 2008 in three counties with a high incidence of MPN were eligible. Cases (n=55) diagnosed between 2001 and 2010 were identified from the Pennsylvania cancer registry and a previous MPN cluster investigation. A panel of physicians verified all diagnoses. Controls (n=473) were recruited through random digit dialling. We collected lifetime, residential, smoking, and dietary histories during telephone interviews. Blood samples for genotyping from 31 cases and 292 controls were obtained.

Results

Cases were older and more likely to be male compared to controls, but otherwise demographically similar. Other lifestyle factors, including residential history and diet associated with exposure to PAH, were not associated with the outcomes. The results were not altered by restricting cases to those with the JAK2 V617F mutation.

Conclusions

We found no evidence of relationships between risk of MPNs and residential history, diet, or lifestyle behaviors associated with exposure to aromatic and heterocyclic amines. However, we did not confirm PAH exposure using sources of information other than the interviews; therefore, exposure misclassification is a limitation of this analysis. Participation rates were poor, raising the possibility that results may be influenced by selections bias. Our results do, in fact, provide evidence that PAHs are an unlikely etiologic agent for MPNs since no association was observed to cigarette smoke (rich in PAHs).

Introduction

Polycythemia vera (PV), essential thrombocythemia (ET) and primary myelofibrosis (PMF) are a category of myeloproliferative neoplasms (MPN) that are characterized by an overproduction of erythrocytes and platelets (Campbell & Green, 2005). They are diseases of the bone marrow, with myeloproliferative manifestation, meaning there is an excess of the otherwise healthy blood cells.

In 2008, the joint efforts by the Pennsylvania Department of Health (PADOH) and the Agency for Toxic Substances and Registry (ATSDR) of the Centers for Disease Control and Prevention (CDC) confirmed a cancer cluster in a tri-county region of Northeast Pennsylvania. These investigations found the incidence of PV to be 4.3 times higher within the cluster area than a comparative tri-county region.

In 2001, PV, ET and PMF were required to be reported to the states cancer registries (Seaman, 2009). However, Seaman et.al found that for persons diagnosed in years 2001-2005, there was still a high rate (59%) of unreported cases which suggests active case finding is still needed to complete cluster investigation (Seaman, 2009). Only hospital cases were required to be reported to the registry, and treatment of MPNs typically does not require hospitalizion. Symptoms that present in PV cases with increased red blood cell production could also be caused by chronic respiratory disease resulting from heavy smoking, renal disease, or chronic pulmonary disease (Stuart & Viera, 2004). This has contributed to the presence of false positives found in cancer registries and even mis-diagnosis of PV prior to the discovery of the JAK2 V617F mutation.

Prior to the 2008 revised diagnostic criterion, in 2001-2004 the National Cancer Institute estimated incidence of PV at 2.8 per 100,000 persons per year (Kutti &Ridell,

2001). For ET the incidence rate was estimated at 1.5 cases per 100,000 per year (Kutti &Ridell, 2001). The rarest of the MPNs is PMF, with an incidence rate of only 0.4 cases per 100,000 persons per year (Kutti &Ridell, 2001). PV patients may be asymptomatic. The diagnosis is now largely based on laboratory tests. Most of the health concerns associated with PV are the result of thickening of the blood, such as clotting. Itching is a classic symptom of PV and usually occurs after taking a bath, or other exposures to warm water (Steinman et al., 1987).

The JAK2 V617F mutation and its role in MPN

JAK2 is an on-and-off switch that regulates bone marrow activity (Baxter, 2005). The JAK2 V617F mutation activates a tyrosine kinase complex in bone marrow normally responsible for regulating blood cell production through molecular signaling. The mutation disrupts the normal inhibition of growth, thus increasing blood cell production (Spivak et al., 2010). The JAK2-acquired point mutation occurs in nearly all PV patients (>95%) and about half of those with ET or PMF, but not in other cancers (Spivak et al., 2010; Kralovics et al., 2007; Tefferi et al., 2007). Following recommendations by the World Health Organization (WHO) in 2008, physicians are recommended to test for the somatic point mutation, JAK2 617V>F (JAK2), to establish diagnosis of MPN (Seaman et al., 2010). This has brought new interest into the etiology of the disease (Anderson et al.,2012).

Polyaromatic hydrocarbons (PAHs)

PAHs are formed from ringed structures of carbon atoms with attached hydrogen atoms, and other atoms as well. PAHs are found in any combustion product, including tobacco smoke, vehicle exhaust, and waste incinerator emissions (Irigaray et al., 2007). PAHs can adhere to fine particulate matter and are in the inhaled air (Irigaray et al., 2007).

This group of compounds is made of two or more fused aromatic rings. PAHs – which are represented by over 100 distinct compounds that may mix together – include chemicals that are known carcinogens in humans (Hayes et al., 1990). Primary environmental sources of PAHs include coal, oil and gas burning, and cigarette smoking. Dietary sources include eating charbroiled or barbecued meats and grains or vegetables contaminated by ambient PAHs (Hayes et al., 1990; Rothman et al., 1993).

PAHs can be a significant part of respirable particulate matter, and are a source of genotoxic activities (Singh et al., 2007; Irigaray et al., 2007). A study of an occupational cohort in Italy reported high PAH exposure in some industries, including coal tar, coke production, and aluminium production, and found an increased risk of bladder and lung cancer in workers with PAH exposure (Bosetti et al., 2007). A study of air-polluted cities found an 8% increase in death risk from lung cancer compared to less polluted cities, even after controlling for smoking (Irigaray et al., 2007). From the literature on PAH exposure and other cancers, it is logical to study PAHs and PV, ET, PMF, and JAK2 V617F.

The suspected association of environmental PAHs with MPNs

PAHs are ubiquitous by-products of organic matter combustion processes and are considered to be mutagens (Hung et al., 2003). The Environmental Protection Agency has only identified sixteen of these PAHs as priority pollutants. Of these, there are as least eight that could be considered as probable or possible human carcinogens, including benzo[a]pyrene, dibenzo[a,h]anthracene, benzo[b]fluoranthene and indenol[1,2,3,c-d]pyrene, benzo[k]fluoranthene (Phillips, 1999). Long-term exposure to

PAHs though air pollution has been associated with increases in morbidity and mortality (Galbraith et al., 2010; Stuart & Viera, 2004; Steinman et al., 1987; Irigaray et al., 2007).

A recent sub-study of the Northern California Childhood Leukemia Study measured PAH exposure in the residential indoor dust in homes of 227 controls and 210 cases with childhood leukemia (Rull et al., 2008). They observed, in single-PAH risk models comparing highest to lowest quartiles, elevated odds ratios for three PAHs: benzo[k]fluoranthene (1.8, 95%CI= 1.0-3.5), benzo[b]fluoranthene (2.0, 95% confidence interval (CI)= 1.0-3.9), and indeno[1,2,3-cd]pyrene (1.7, 95%CI= 0.9-3.2) (Rull et al., 2008). This provides support that changes in bone marrow resulting from a chemical exposure may have a role in chronic leukemia diseases like MPNs. The study also highlighted a potential mechanism for the route of exposure to PAHs that cling to air particles of greater than $2.5u.m$ (Rull et al., 2008). Respirable PAHs come from many sources, but most commonly from point sources, like power plants, residential heating, and mobile sources like motor vehicles (Phillips, 1999; Steinman et al., 1987).

Pasqualetti, et al. (1991) reported an increased risk of 2.15 (95% CI=1.39, 3.32), with self-reported exposure to aromatic hydrocarbons for all hematological malignances in a hospital-based, case-control study in Italy (Pasqualettiet al., 1991). Among the 620 cases, there were 44 with myeloproliferative syndromes, with 576 cases having one of the following malignancies (Pasqualetti et al., 1991): acute non-lymphoblastic leukemia; non-Hodgkin's lymphoma; Hodgkin's disease; myelodysplastic syndromes; malignant monoclonal gammopathies; multiple myeloma and chronic lymphocytic leukemia.

However, the study did not include analysis for myeloproliferative syndromes separately in the published report (Pasqualetti et al., 1991).

The focus of this analysis is PAHs. Since we could offer no strong a priori hypothesis about specific exposure sources, we consider a range of potential avenues of PAH exposure, especially those previously implicated in blood tumors, in our investigation. We expect that environmental PAHs from residential history, diet, lifestyle behaviors with presumed exposures might show associations with MPNs. Although chemicals like PAHs are everywhere, they do represent a burdensome chemical exposure category for the PA geographic region containing the PV cluster given the extensive coal mining activities and the number of waste coal power plants in the tri-county area (Seaman et al., 2009).

Methods

The project used a case-control design. The study area was comprised of a tri-county area in Northeast Pennsylvania (Carbon, Luzerne, and Schuylkill Counties). Cases included individuals with diagnoses of PV, ET, and PMF, with and without the JAK2 V617F mutation. Study of prevalent cases is justified because mortality associated with these conditions is low (0.002 to 2.8 per 100,000 persons (Kutti & Ridell, 2001), so changes in prevalence will primarily reflect changes in incidence. There is limited opportunity for incidence-prevalence bias, so prevalence odds ratios estimated in this design closely approximated prevalence rate ratios.

Control selection

We selected a stratified random sample of controls from the study area. The sample collected represents the population distribution based on the 2006-2008 American Community Survey Population Estimates for: age (60% 42-59, 40% 60-89); gender (50% male, 50% female); county of residence (15% Carbon, 55% Luzerne, and 30% Schuylkill County. Control subjects were drawn from a purchased published residence lists for the tri-county area. We contacted current residents at random by phone to seek eligible persons interested in receiving the study mailing and consent forms. We asked subjects who consented to receive mailings to return signed consent forms. Those sent, but not returned consents received a follow-up phone call where the research team sought their oral consent. We continued to contact residents until the tri-county population was sufficiently sampled and the controls reflected the source population.

Data collection

The Geisinger Survey Research Unit interviewed all consenting participants. The research unit's staff administered the questionnaire in a standardized manner using a computer-assisted telephone interview. Interviewers were initially blinded to case or control status of participants until the end of the interview, when they were asked about individual and familial history of blood diseases. The questionnaire elicited detailed information about demographic characteristics, health behaviors, socioeconomic information, residential history, and past experience with conditions associated with exposure to hazardous materials (full questionnaire and IRB documents are in the Appendices A to E).

Consenting participants received a summary of the residential and occupational history questions in advance of the interview to help prepare them. The interviews lasted between 45 minutes and about two hours. We mailed a gift certificate incentive of $25 to participants after completion of the interview. We also mailed and phoned eligible cases that had relocated since 2008 from the tri-county area. We only conducted case interviews to surviving patients that consented to participate.

Lifestyle exposures

Lifestyle exposures were collected from multiple exposure sources. This included questions on smoking history and diet. The dietary questions used in our study were adapted from the Long Island Breast Cancer Study Questionnaire, available online through the National Cancer Institute website. We collected consumption of grilled, barbequed, or smoked meat or fish by decade by five categorizations of different meats types per season, including; all year, summer, fall, winter and spring.

Residential exposures

Each subject was asked to list every primary residence they had lived in for six months or longer, beginning with current residence and working backwards, listing all residences back to age 21 with a maximum of nine. We assessed distance to different industry types as well as hazardous waste sites and waste coal power plants.

Analytic method

We conducted descriptive analysis on the characteristics of the study population. Logistic regression was used to estimate crude and adjusted odds ratios (OR) and associated 95% confidence intervals (CI) for each individual exposure measure with no or minimal exposure as the reference. Adjusted models included age, sex, and county as covariates.

Results

Only 27% of cases consented for the telephone interview and only 56 % of those also consented to the optional blood draw. Response rate among our controls was slightly better, at 41% (which is fair for a Radom Digit Dialling protocol to consent controls), with 61% of those also consenting for the blood draw.

Demographics

Cases were older (median age=71 vs. 61yrs) (OR=3.0, 95% CI=1.5, 5.8) and more likely to be male (49% vs. 39%) (OR=1.5, 95% CI=0.8, 2.8) compared to controls, but they were otherwise demographically similar (Table 2.3). The study population was overwhelmingly Caucasian (98% of cases and 99% of controls). None of the cases and few (2%) controls were of Jewish ancestry and all participants except for one control were born in the US. Two-thirds of the subjects were married at the time of the interview. More cases than controls were retired (63% vs. 42%). These findings were similar across all case categorizations (Table 2.4).

Diet history

Most of the cases (93%) and controls (97%) reported some exposure to grilled, barbequed, or smoked meat or fish (see Table 2.5). In the analysis of the five

categorization per season of consumption of grilled, barbecued, or smoked meat or fish and seasonality, there were only three associations that reach statistical significance with having an MPN. In the adjusted logistic regression models, we found that JAK2 V617F cases were more likely than controls to eat grilled or barbequed poultry in the winter (aOR =3.5, 95% CI=1.2,10.0) (see Table 2.6). However, there was no significant differences between our cases and controls for eating eat grilled or barbequed poultry in the fall or the summer. All cases ate grilled or barbequed poultry or fish in the summer, and for the fall it was approximately 50% across all case categorizations. In addition, cases were less likely to have ever eaten grilled or barbequed poultry or fish, significant across all case categorizations (aOR=0.3 -0.4, 95% CI=0.1, 0.9) and for JAK2 V617F cases (aOR=0.4, 95% CI=0.1, 0.9). There were also no significant differences between cases and controls for eating fish from Still Creek or other local lakes or creeks (see Tables 2.9 and 2.10). We also found no relationships between MPNs and residential history (Table 2.7 and 2.8).

Smoking

Half the controls ever smoked, as did 32%-45% of cases in Table 2.11. Out of the 49 % of cases who ever smoked, all (100%) smoked cigarettes and only eight confirmed JAK2 V617F cases ever smoked (32% , 95% CI:14,50). Table 2.12 shows the results for the effect of smoking by pack years as well as smoked filtered vs. non-filtered cigarettes. Cases were less likely to be heavy smokers (>15 pack years) compared to non-smokers, with an adjusted OR of 0.2 (95% CI: <0.0, 0.8) for JAK2 V617F cases. Results were similar although not statistically significant when heavy smokers were compared to light smoker s (<15 pack years), with an adjusted OR of 0.2 (95% CI= <0.0, 1.4) for JAK2 V617F cases.

Discussion

To our knowledge, this is the first study to look at PAH exposure from diet and smoking and risk of developing an MPN.

We did not see any evidence of associations between PAH exposures and MPNs

Significant lifestyle sources of PAH exposure are cigarette smoking and eating well-barbequed meat and vegetables (Hayes, 1990; Rothman, 1993). We saw a significant increased risk of MPNs in subjects who ate grilled or barbecued poultry or fish all year long (aOR 3.5, 95% CI: 1.2, 10.0), but a protective effect from ever eating grilled or barbecued poultry or fish (aOR 0.3, 95% CI: 0.2, 0.6). All 38 other comparisons of significant exposure sources hovered around the null and were similar for all case categorizations, including JAK2- cases reporting no associations.

Results for our analysis with smoking consistently suggested no association. If PAHs were associated with an increased risk of developing an MPN, we would expect to see an association when examining smoking status, perhaps at least as strong (if not stronger), than that of eating grilled or barbequed poultry or fish all year long (high exposure to PAHs through diet) (Hayes et al., 1990; Rothman et al., 1993). Our study had a detailed self-reported smoking history and since MPNs haven't previously been thought to be associated with smoking, we do not expect differential misclassification of self-reported smoking status in cases or controls in this older population.

Our findings do not support an association of PAHs and increased risk of getting an MPN. However, this study was only had the power to detect very large effects (>4 ORs), so any small effects from PAHs would not be detected as statistically significant.

Limitations and strengths of the study

This study has several limitations. The first limitation of this study is the number of cases that were recruited from the area was only 55. We did use state-of-the-art diagnostic criteria not used in past case-control studies and had cases reviewed by an independent expert panel. However, the changing of the WHO diagnostic criteria (changed in both 2001 and 2008) was a challenge. In 2001 PV, ET, and PMF became a reportable cancer under the ICD-O codes, and the WHO started reclassifying these disorders based on histological and molecular information (Vakil & Tefferi, 2011). This included adding a molecular test (JAK2 V617F mutation) as diagnostic criteria in 2008 and adding four other MPN–unclassifiable diseases into the MPN disease category (Vakil & Tefferi, 2011). The ability to test cases for the JAK2 V617F mutation directly did attempt to minimize misclassification of our cases.

Our study also suffered from a small response rate from our MPN cases. The optional blood draw was a critical step in our study to help counter the drift in diagnostic criteria for this group on MPNs, especially PV. Since each case must consent to participate, this was a moving target that determined the total number of cases in our study in a short period of time. As with all retrospective studies, there was a potential for recall bias.

A large limitation of this study is the potential for bias from misclassification of exposure. We did not confirm PAH exposure with sources of information other than the interviews. Participation rates were poor, particularly for cases, which raise concerns about selection bias; however, our use of random selection ensured the controls were representative of the tri county area population.

We were able to measure PAH exposures from multiple sources, including residence and lifestyle, and included a significant source to non-smokers (diet). Taken together, the results do not support the role of PAH in etiology of MPNs.

List of References

Anderson, L. A., Duncombe, A. S., Hughes, M., Mills, M. E., Wilson, J. C., & McMullin, M. F. (2012). Environmental, lifestyle, and familial/ethnic factors associated with myeloproliferative neoplasms. *Am J Hematol, 87*(2), 175-182.

Baxter EJ,Scott LM,Campbell PJ, et al. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet* 2005; **365**: 1054–1061.

Buchanich, J., Mertz, K., 2013. Updated and Expanded Study of Polycythemia Vera and Other Myeloproliferative Neoplasms in the Tri-County Area. FINAL REPORT to Pennsylvania Department of Health

Campbell, P. J., & Green, A. R. (2005). Management of polycythemia vera and essential thrombocythemia. *Hematology Am Soc Hematol Educ Program*, 201-208.

Clapp, R. W., Jacobs, M. M., & Loechler, E. L. (2008). Environmental and occupational causes of cancer: new evidence 2005-2007. *Rev Environ Health, 23*(1), 1-37.

Galbraith, D., Gross, S. A., & Paustenbach, D. (2010). Benzene and human health: A historical review and appraisal of associations with various diseases. *Crit Rev Toxicol, 40 Suppl 2*, 1-46.

Hayes, R. B., Blair, A., Stewart, P. A., Herrick, R. F., & Mahar, H. (1990). Mortality of U.S. embalmers and funeral directors. *Am J Ind Med, 18*(6), 641-652.

Hung, R. J., Boffetta, P., Brockmoller, J., Butkiewicz, D., Cascorbi, I., Clapper, M. L., et al. (2003). CYP1A1 and GSTM1 genetic polymorphisms and lung cancer risk in Caucasian non-smokers: a pooled analysis. *Carcinogenesis, 24*(5), 875-882.

Irigaray, P., Newby, J. A., Clapp, R., Hardell, L., Howard, V., Montagnier, L., .Belpomme, D. (2007). Lifestyle-related factors and environmental agents causing cancer: an overview. *Biomed Pharmacother, 61*(10), 640-658.

Kralovics R, Passamonti F, Buser AS, Teo S, Tiedt R, Passweg JR, Tichelli A, Cazzola M, Skoda RC. A gain-of-function mutation of JAK2 in myeloproliferative disorders. N. Engl. J. Med. 2005;352:1779–1790.

Kutti, J., & Ridell, B. (2001). Epidemiology of the myeloproliferative disorders: essential thrombocythaemia, polycythaemia vera and idiopathic myelofibrosis. *Pathol Biol (Paris), 49*(2), 164-166.

Occupational exposures in petroleum refining; crude oil and major petroleum fuels. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. (1989). *IARC Monogr Eval Carcinog Risks Hum, 45*, 1-322.

Pasqualetti, P., Casale, R., Colantonio, D., & Collacciani, A. (1991). Occupational risk for hematological malignancies. *Am J Hematol, 38*(2), 147-149.

Passamonti, Francesco, & Rumi, Elisa. (2009). Clinical relevance of JAK2 (V617F) mutant allele burden. *Haematologica, 94*(1), 7-10.

Phillips, D. H. (1999). Polycyclic aromatic hydrocarbons in the diet. *Mutat Res, 443*(1-2), 139-147.

Purdue, M. P., Hoppin, J. A., Blair, A., Dosemeci, M., & Alavanja, M. C. (2007). Occupational exposure to organochlorine insecticides and cancer incidence in the Agricultural Health Study. *Int J Cancer, 120*(3), 642-649.

Rothman, N., Correa-Villasenor, A., Ford, D. P., Poirier, M. C., Haas, R., Hansen, J. A., et al. (1993). Contribution of occupation and diet to white blood cell polycyclic aromatic hydrocarbon-DNA adducts in wildland firefighters. *Cancer Epidemiol Biomarkers Prev, 2*(4), 341-347.

Rull, R P, Gunier, R B, Reynolds, P, Colt, J S, Nishioka, M, Metayer, C, Ward, M H. (2008). Residential Concentrations of Polycyclic Aromatic Hydrocarbons and Childhood Leukemia Risk. *Epidemiology, 19*(6), S275

Seaman, V., Jumaan, A., Yanni, E., Lewis, B., Neyer, J., Roda, P., Hoffman, R. (2009). Use of molecular testing to identify a cluster of patients with polycythemia vera in eastern Pennsylvania. *Cancer Epidemiol Biomarkers Prev, 18*(2), 534-540.

Singh, Rajinder, Kaur, Balvinder, Kalina, Ivan, Popov, Todor A., Georgieva, Tzveta, Garte, Seymour, Farmer, Peter B. (2007). Effects of environmental air pollution on endogenous oxidative DNA damage in humans. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 620*(1-2), 71-82.

Spivak JL. (2010) Narrative review: Thrombocytosis, polycythemia vera, and JAK2 mutations:The phenotypic mimicry of chronic myelofibrosis. Ann Intern Med. 2010 Mar 2;152(5):300-6. Review.

Steinman, H. K., Kobza-Black, A., Lotti, T. M., Brunetti, L., Panconesi, E., & Greaves, M. W. (1987). Polycythaemia rubra vera and water-induced pruritus: blood histamine levels and cutaneous fibrinolytic activity before and after water challenge. *Br J Dermatol, 116*(3), 329-333.

Stuart, B. J., & Viera, A. J. (2004). Polycythemia vera. *Am Fam Physician, 69*(9), 2139-2144.

Tefferi, A., Thiele, J., Orazi, A., Kvasnicka, H. M., Barbui, T., Hanson, C. A., Vardiman, J. W. (2007). Proposals and rationale for revision of the World Health Organization diagnostic criteria for polycythemia vera, essential thrombocythemia, and primary myelofibrosis: recommendations from an ad hoc international expert panel. *Blood, 110*(4), 1092-1097.

Vakil, E., & Tefferi, A. (2011). BCR-ABL1--negative myeloproliferative neoplasms: a review of molecular biology, diagnosis, and treatment. *Clin Lymphoma Myeloma Leuk, 11 Suppl 1*, S37-45.

Table 2.1 The demographics of the tri-county region (Carbon, Luzerne, and Schuylkill Counties) in Northeast Pennsylvania

County	Population [1]	Percentage of Population Caucasian	Median Age [1]
Carbon	62,937	96.1	41.4
Luzerne	311,752	94.4	42.1
Schuylkill	147,107	95.3	42.0
Total Tri-County	521,796	94.9	42.0

1. Data retrieved from U.S. Census American Community Survey, 2008. For 1921-1967 DOBs, we looked for ages 33-79 for the year 2000, ages 40-86 for 2007, and ages 43-89 for 2010. Census data age group numbers do not exactly match the 1921-1968 DOB cohort due to differences in intervals

Table 2.2 Case Categories

Consenting Case Subgroups [1]	Total Interviewed	With Genotype Data
All cases	55	31
Confirmed cases	41	27
Confirmed PV cases	33	24
Confirmed ET cases	7	2
Confirmed PMF cases	1	0
JAK2 V617F mutation with PV, ET or PMF	25	22

1.All cases include PV, ET and PMF, confirmed cases by expert panel include PV, ET, and PMF, confirmed JAK2 V617F mutation with PV, ET or PMF by expert panel, confirmed by expert panel PV only

Table 2.3 Study demographics [1]

	All cases [2] n= 55	Confirmed cases [2] n= 41	Confirmed JAK2 V617Fcases [2] n= 25	Confirmed PV cases [2] n= 33	Controls n= 473
County					
Carbon	7 (13%)	6 (15%)	4 (16%)	6 (18%)	77 (16%)
Luzerne	27 (49%)	20 (49%)	13 (52%)	13 (39%)	267 (56%)
Schuylkill	21 (38%)	15 (37%)	8 (32%)	14 (42%)	129 (27%)
Age					
42-64	19 (35%)	14 (34%)	7 (28%)	10 (30%)	287 (61%)
65+	36 (65%)	27 (66%)	18 (72%)	23 (70%)	186 (39%)
Sex					
Male	27 (49%)	20 (49%)	12 (48%)	18 (55%)	185 (39%)
Female	28 (51%)	21 (51%)	13 (52%)	15 (45%)	288 (61%)
Race/ethnicity					
Non-Hisp. White	54 (98%)	40 (98%)	25 (100%)	33 (100%)	468 (99%)
Latino, Hispanic	0	0	0	0	5 (1%)
Non-Hisp. Black	0	0	0	0	0
Non-Hisp. Native American, Alaskan native	0	0	0	0	0
Non-Hisp. Asian or Pacific Islander	1 (2%)	1 (2%)	0	0	0
Non-Hisp. Other	0	0	0	0	0
Non-Hisp., multiple race	0	0	0	0	0
Country of origin					
USA	55 (100%)	41 (100%)	25 (100%)	33 (100%)	472 (100%)
Other	0	0	0	0	1 (0%) [2]

Table 2.3 Study demographics [1] (cont'd)

	All cases [2] n= 55	Confirmed cases [2] n= 41	Confirmed JAK2 V617Fcases [2] n= 25	Confirmed PV cases [2] n= 33	Controls n= 473
State born in					
AK	0	0	0	0	1 (0%)
CA	1 (2%)	1 (2%)	0	0	4 (1%)
CT	1 (2%)	1 (2%)	1 (4%)	1 (3%)	0
FL	0	0	0	0	1 (0%)
IL	0	0	0	0	1 (0%)
MA	1 (2%)	1 (2%)	1 (4%)	1 (3%)	1 (0%)
MN	0	0	0	0	2 (0%)
MS	0	0	0	0	1 (0%)
NC	0	0	0	0	1 (0%)
NH	0	0	0	0	1 (0%)
NJ	2 (4%)	2 (5%)	2 (8%)	2 (6%)	14 (3%)
NM	0	0	0	0	1 (0%)
NY	0	0	0	0	16 (3%)
OH	0	0	0	0	2 (0%)
OK	1 (2%)	0	0	0	0
PA	48 (87%)	35 (85%)	20 (80%)	28 (85%)	420 (89%)
SC	0	0	0	0	1 (0%)
SD	0	0	0	0	1 (0%)
TN	0	0	0	0	1 (0%)
TX	0	0	0	0	1 (0%)
VA	0	0	0	0	2 (0%)
WV	1 (2%)	1 (2%)	1 (4%)	1 (3%)	0
non-US	0	0	0	0	1 (0%)

Table 2.3 Study demographics [1] (cont'd)

	All cases [2] n= 55	Confirmed cases [2] n= 41	Confirmed JAK2 V617Fcases [2] n= 25	Confirmed PV cases [2] n= 33	Controls n= 473
Jewish ancestry					
Yes	0	0	0	0	10 2%
No	54 (100%)	40 (100%)	25 (100%)	33 (100%)	455 (98%)
Don't know	0	0	0	0	5
Missing	1	1	0	0	3
Marital status					
Married	34 (62%)	27 (66%)	16 (64%)	21 (64%)	295 (62%)
Widowed	13 (24%)	11 (27%)	7 (28%)	9 (27%)	76 (16%)
Currently single	8 (15%)	3 (7%)	2 (8%)	3 (9%)	102 (22%)
Education					
Less than high school	5 (9%)	3 (7%)	2 (8%)	2 (6%)	11 (2%)
High school / GED	30 (55%)	22 (54%)	13 (52%)	20 (61%)	180 (38%)
Some college	7 (13%)	7 (17%)	5 (20%)	5 (15%)	159 (34%)
Bachelors degree	6 (11%	3 (7%)	2 (8%)	2 (6%)	58 (12%)
More than bachelors	7 (13%)	6 (15%)	3 (12%)	4 (12%)	65 (14%)
Household income					
Less than $20,000	8 (16%)	5 (14%)	3 (14%)	5 (17%)	67 (15%)
$20,000 - $35,000	14 (28%)	10 (27%)	7 (32%)	10 (33%)	106 (24%)
$35,000 - $50,000	8 (16%)	5 (14%)	2 (9%)	4 (13%)	91 (20%)
$50,000 - $75,000	10 (20%)	9 (24%)	4 (18%)	8 (27%)	97 (22%)
More than $75,000	10 (20%)	8 (22%)	6 (27%)	3 (10%)	83 (19%)
Don't know	3	3	2	2	8
Refused	2	1	1	1	21

Table 2.3 Study demographics [1] (cont'd)

	All cases [2] n= 55	Confirmed cases [2] n= 41	Confirmed JAK2 V617Fcases [2] n= 25	Confirmed PV cases [2] n= 33	Controls n= 473
Current employment					
Employed for wages	17 (31%	12 (29%)	7 (28%)	9 (27%)	184 (39%)
Self-employed	1 (2%)	1 (2%)	1 (4%)	1 (3%)	16 (3%)
Out of work for more than a year	2 (4%)	1 (2%)	1 (4%)	0	15 (3%)
Out of work for less than a year	0	0	0	0	10 (2%)
Homemaker	0	0	0	0	15 (3%)
Student	0	0	0	0	1 (0%)
Retired	32 (58%)	26 (63%)	16 (64%)	22 (67%)	200 (42%)
Unable to work	3 (5%)	1 (2%)	0	1 (3%)	32 (7%)

1. Self-reported socio economic factors, number and %
2. All cases include PV, ET and PMF, confirmed cases by expert panel include PV, ET, and PMF, confirmed JAK2 V617F mutation with PV, ET or PMF by expert panel, confirmed by expert panel PV only

Table 2.4 Study demographic logistic regression models [1]

	All cases OR [2] (95% CI) n=55	All cases aOR [3] (95% CI) n=55	Confirmed cases OR [2] (95% CI) n=41	Confirmed cases aOR [3] (95% CI) n=41	Confirmed JAK2 V617F OR [2] (95% CI) n=25	Confirmed JAK2 V617F aOR [3] (95% CI) n=25	Confirmed PV OR [2] (95% CI) n=33	Confirmed PV aOR [3] (95% CI) n=33
County								
Carbon	0.9 (0.4, 2.1)	0.9 (0.4, 2.1)	1.0 (0.4, 2.7)	1.0 (0.4, 2.6)	1.1 (0.3, 3.4)	1.0 (0.3, 3.3)	1.6 (0.6, 4.4)	1.6 (0.6, 4.3)
Luzerne	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Schuylkill	1.6 (0.9, 3.0)	1.5 (0.8, 2.8)	1.6 (0.8, 3.1)	1.4 (0.7, 2.9)	1.3 (0.5, 3.2)	1.2 (0.5, 2.9)	2.2 (1.0, 4.9)	2.0 (0.9, 4.4)
Age								
42-64		Referent		Referent		Referent		Referent
65+		2.9 (1.6, 5.2)		3.0 (1.5, 5.8)		4.0 (1.6, 9.7)		3.5 (1.6, 7.5)
Sex								
Male		1.4 (0.8, 2.5)		1.5 (0.8, 2.8)		1.4 (0.6, 3.2)		1.8 (0.9, 3.7)
Female		Referent		Referent		Referent		Referent
Age								
42-64	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
65+	2.9 (1.6, 5.3)	2.9 (1.6, 5.2)	3.0 (1.5, 5.8)	3.0 (1.5, 5.8)	4.0 (1.6, 9.7)	4.0 (1.6, 9.7)	3.5 (1.7, 7.6)	3.5 (1.6, 7.5)
Sex								
Male		1.4 (0.8, 2.5)		1.5 (0.8, 2.8)		1.4 (0.6, 3.2)		1.8 (0.9, 3.7)
Female		Referent		Referent		Referent		Referent

Table 2.4 Study demographic logistic regression models [1] (cont'd)

	All cases OR [2] (95% CI) n=55	All cases aOR [3] (95% CI) n=55	Confirmed cases OR [2] (95% CI) n=41	Confirmed cases aOR [3] (95% CI) n=41	Confirmed JAK2 V617F OR [2] (95% CI) n=25	Confirmed JAK2 V617F aOR [3] (95% CI) n=25	Confirmed PV OR [2] (95% CI) n=33	Confirmed PV aOR [3] (95% CI) n=33
County								
Carbon		0.9 (0.4, 2.1)		1.0 (0.4, 2.6)		1.0 (0.3, 3.3)		1.6 (0.6, 4.3)
Schuylkill		1.5 (0.8, 2.8)		1.4 (0.7, 2.9)		1.2 (0.5, 2.9)		2.0 (0.9, 4.4)
Luzerne		Referent		Referent		Referent		Referent
Sex								
Male	1.5 (0.9, 2.6)	1.4 (0.8, 2.5)	1.5 (0.8, 2.8)	1.5 (0.8, 2.8)	1.4 (0.6, 3.2)	1.4 (0.6, 3.2)	1.9 (0.9, 3.8)	1.8 (0.9, 3.7)
Female	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Age								
42-64		Referent		Referent		Referent		Referent
65+		2.9 (1.6, 5.2)		3.0 (1.5, 5.8)		4.0 (1.6, 9.7)		3.5 (1.6, 7.5)
County								
Carbon		0.9 (0.4, 2.1)		1.0 (0.4, 2.6)		1.0 (0.3, 3.3)		1.6 (0.6, 4.3)
Schuylkill		1.5 (0.8, 2.8)		1.4 (0.7, 2.9)		1.2 (0.5, 2.9)		2.0 (0.9, 4.4)
Luzerne		Referent		Referent		Referent		Referent

Table 2.4 Study demographic logistic regression models [1] (cont'd)

	All cases OR [2] (95% CI) n=55	All cases aOR [3] (95% CI) n=55	Confirmed cases OR [2] (95% CI) n=41	Confirmed cases aOR [3] (95% CI) n=41	Confirmed JAK2 V617F OR [2] (95% CI) n=25	Confirmed JAK2 V617F aOR [3] (95% CI) n=25	Confirmed PV OR [2] (95% CI) n=33	Confirmed PV aOR [3] (95% CI) n=33
Marital status								
Married	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Widowed	1.5 (0.7, 3.0)	1.1 (0.5, 2.2)	1.6 (0.8, 3.3)	1.1 (0.5, 2.6)	1.7 (0.7, 4.3)	1.1 (0.4, 3.0)	1.7 (0.7, 3.8)	1.2 (0.5, 3.1)
Currently single	0.7 (0.3, 1.5)	0.7 (0.3, 1.6)	0.3 (0.1, 1.1)	0.3 (0.1, 1.1)	0.4 (0.1, 1.6)	0.4 (0.1, 1.7)	0.4 (0.1, 1.4)	0.4 (0.1, 1.6)
Age								
42-64		Referent		Referent		Referent		Referent
65+		2.8 (1.5, 5.2)		2.6 (1.3, 5.4)		3.6 (1.4, 9.2)		3.0 (1.3, 6.9)
Sex								
Male		1.4 (0.8, 2.6)		1.4 (0.7, 2.8)		1.4 (0.6, 3.3)		1.8 (0.8, 3.8)
Female		Referent		Referent		Referent		Referent
County								
Carbon		0.9 (0.4, 2.1)		1.0 (0.4, 2.6)		1.0 (0.3, 3.2)		1.5 (0.6, 4.2)
Schuylkill		1.6 (0.8, 2.9)		1.5 (0.7, 3.2)		1.3 (0.5, 3.2)		2.1 (0.9, 4.8)
Luzerne		Referent		Referent		Referent		Referent

Table 2.4 Study demographic logistic regression models [1] (cont'd)

	All cases OR [2] (95% CI) n=55	All cases aOR [3] (95% CI) n=55	Confirmed cases OR [2] (95% CI) n=41	Confirmed cases aOR [3] (95% CI) n=41	Confirmed JAK2 V617F OR [2] (95% CI) n=25	Confirmed JAK2 V617F aOR [3] (95% CI) n=25	Confirmed PV OR [2] (95% CI) n=33	Confirmed PV aOR [3] (95% CI) n=33
Education								
Less than high school	4.2 (1.1, 15.7)	2.8 (0.7, 10.9)	3.0 (0.6, 13.6)	1.9 (0.4, 9.3)	3.9 (0.6, 26.3)	2.4 (0.3, 17.4)	3.0 (0.5, 18.1)	1.8 (0.3, 12.1)
High school / GED	1.5 (0.6, 3.7)	1.2 (0.5, 3.0)	1.3 (0.5, 3.4)	1.0 (0.4, 2.7)	1.6 (0.4, 5.7)	1.2 (0.3, 4.3)	1.8 (0.6, 5.5)	1.4 (0.4, 4.4)
Some college	0.4 (0.1, 1.2)	0.4 (0.1, 1.1)	0.5 (0.2, 1.5)	0.4 (0.1, 1.3)	0.7 (0.2, 2.9)	0.6 (0.1, 2.5)	0.5 (0.1, 2.0)	0.4 (0.1, 1.7)
Bachelors degree	1.0 (0.3, 3.0)	0.9 (0.3, 2.9)	0.6 (0.1, 2.3)	0.6 (0.1, 2.4)	0.7 (0.1, 4.6)	0.7 (0.1, 4.7)	0.6 (0.1, 3.2)	0.6 (0.1, 3.3)
More than bachelors	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Age								
42-64		Referent		Referent		Referent		Referent
65+		2.7 (1.5, 4.9)		2.7 (1.4, 5.5)		3.6 (1.5, 9.1)		3.1 (1.4, 6.9)
Sex								
Male		1.4 (0.8, 2.5)		1.4 (0.7, 2.7)		1.4 (0.6, 3.2)		1.8 (0.9, 3.7)
Female		Referent		Referent		Referent		Referent

Table 2.4 Study demographic logistic regression models [1] (cont'd)

	All cases OR [2] (95% CI) n=55	All cases aOR [3] (95% CI) n=55	Confirmed cases OR [2] (95% CI) n=41	Confirmed cases aOR [3] (95% CI) n=41	Confirmed JAK2 V617F OR [2] (95% CI) n=25	Confirmed JAK2 V617F aOR [3] (95% CI) n=25	Confirmed PV OR [2] (95% CI) n=33	Confirmed PV aOR [3] (95% CI) n=33
County								
Carbon		0.9 (0.4, 2.2)		1.0 (0.4, 2.6)		1.0 (0.3, 3.4)		1.5 (0.5, 4.2)
Schuylkill		1.5 (0.8, 2.8)		1.4 (0.7, 2.9)		1.1 (0.5, 2.9)		1.9 (0.9, 4.3)
Luzerne		Referent		Referent		Referent		Referent
Household income								
Less than $20,000	1.0 (0.4, 2.7)	0.5 (0.2, 1.4)	0.8 (0.2, 2.5)	0.3 (0.1, 1.3)	0.6 (0.1, 2.6)	0.2 (0.1, 1.2)	2.1 (0.5, 9.0)	1.0 (0.2, 4.7)
$20,000 - $35,000	1.1 (0.5, 2.6)	0.6 (0.2, 1.5)	1.0 (0.4, 2.6)	0.5 (0.2, 1.5)	0.9 (0.3, 2.8)	0.4 (0.1, 1.4)	2.6 (0.7, 9.8)	1.3 (0.3, 5.5)
$35,000 - $50,000	0.7 (0.3, 1.9)	0.4 (0.2, 1.2)	0.6 (0.2, 1.8)	0.3 (0.1, 1.1)	0.3 (0.1, 1.5)	0.1 (0.0, 0.8)	1.2 (0.3, 5.6)	0.7 (0.1, 3.5)
$50,000 - $75,000	0.9 (0.3, 2.2)	0.7 (0.3, 1.8)	1.0 (0.4, 2.6)	0.8 (0.3, 2.2)	0.6 (0.2, 2.1)	0.4 (0.1, 1.7)	2.3 (0.6, 8.9)	1.7 (0.4, 7.0)
More than $75,000	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Age								
42-64		Referent		Referent		Referent		Referent
65+		3.6 (1.8, 7.1)		3.7 (1.7, 8.1)		5.3 (1.9, 15.2)		3.5 (1.5, 8.3)

Table 2.4 Study demographic logistic regression models [1] (cont'd)

	All cases OR [2] (95% CI) n=55	All cases aOR [3] (95% CI) n=55	Confirmed cases OR [2] (95% CI) n=41	Confirmed cases aOR [3] (95% CI) n=41	Confirmed JAK2 V617F OR [2] (95% CI) n=25	Confirmed JAK2 V617F aOR [3] (95% CI) n=25	Confirmed PV OR [2] (95% CI) n=33	Confirmed PV aOR [3] (95% CI) n=33
Sex								
Male		1.5 (0.8, 2.7)		1.4 (0.7, 2.8)		1.4 (0.6, 3.4)		1.7 (0.8, 3.8)
Female		Referent		Referent		Referent		Referent
County								
Carbon		1.1 (0.4, 2.6)		1.2 (0.5, 3.3)		1.3 (0.4, 4.4)		1.8 (0.6, 5.1)
Schuylkill		1.8 (0.9, 3.5)		1.7 (0.8, 3.6)		1.4 (0.5, 3.8)		2.1 (0.9, 4.9)
Luzerne		Referent		Referent		Referent		Referent

1. Self-reported socio economic factors
2. All cases include PV, ET and PMF, confirmed cases by expert panel include PV, ET, and PMF, confirmed JAK2 V617F mutation with PV, ET or PMF by expert panel, confirmed by expert panel PV only.
3. Adjusted (for age, sex and county) ORs (95% confidence intervals)

Table 2.5. Cooked meat diet history [1]

	All cases [2] n=55	Confirmed cases [2] n=41	Confirmed JAK2 V617Fcases[2] n=25	Confirmed PV cases [2] n=33	Controls n=473
Ever ate grilled, BBQ, or smoked meat or fish					
Yes	52 (95%)	38 (93%)	25 (100%)	31 (94%)	461 (97%)
No	3 (5%)	3 (7%)	0	2 (6%)	12 (3%)
Ate grilled, BBQ, or smoked meat or fish in the winter					
No	34 (65%)	25 (66%)	17 (68%)	22 (71%)	319 (69%)
Yes	18 (35%)	13 (34%)	8 (32%)	9 (29%)	142 (31%)
Missing	3	3	0	2	12
Ate grilled, BBQ, smoked meat or fish in the spring					
No	32 (62%)	23 (61%)	14 (56%)	20 (65%)	250 (54%)
Yes	20 (38%)	15 (39%)	11 (44%)	11 (35%)	211 (46%)
Missing	3	3	0	2	12
Ate grilled, BBQ, smoked meat or fish in the summer					
No	0	0	0	0	5 (1%)
Yes	52 (100%)	38 (100%)	25 (100%)	31 (100%)	456 (99%)
Missing	3	3	0	2	12
Ate grilled, BBQ, smoked meat or fish in the fall					
No	33 (63%)	24 (63%)	15 (60%)	21 (68%)	269 (58%)
Yes	19 (37%)	14 (37%)	10 (40%)	10 (32%)	192 (42%)
Missing	3	3	0	2	12

Table 2.5. Cooked meat diet history[1] (cont'd)

	All cases [2] n=55	Confirmed cases [2] n=41	Confirmed JAK2 V617Fcases[2] n=25	Confirmed PV cases [2] n=33	Controls n=473
Ate grilled, BBQ, smoked meat or fish all year					
No	35 (67%)	26 (68%)	17 (68%)	23 (74%)	323 (70%)
Yes	17 (33%)	12 (32%)	8 (32%)	8 (26%)	138 (30%)
Missing	3	3	0	2	12
Ever ate grilled or BBQ beef, lamb, or pork					
Yes	45 (82%)	35 (85%)	23 (92%)	31 (94%)	421 (89%)
No	10 (18%)	6 (15%)	2 (8%)	2 (6%)	52 (11%)
Ate grilled or BBQ beef, lamb, or pork in the spring					
No	24 (53%)	20 (57%)	12 (52%)	18 (58%)	229 (54%)
Yes	21 (47%)	15 (43 %)	11 (48%)	13 (42%)	192 (46%)
Missing	10	6	2	2	52
Ate grilled or BBQ beef, lamb, or pork in the summer					
No	0	0	0	0	2 (0%)
Yes	45 (100%)	35 (100%)	23 (100%)	31 (100%)	419 (100%)
Missing	10	6	2	2	52
Ate grilled or BBQ beef, lamb, or pork in the fall					
No	26 (58%)	21 (60%)	13 (57%)	19 (61%)	243 (58%)
Yes	19 (42%)	14 (40%)	10 (43%)	12 (39%)	178 (42%)
Missing	10	6	2	2	52

Table 2.5. Cooked meat diet history[1] (cont'd)

	All cases [2] n=55	Confirmed cases [2] n=41	Confirmed JAK2 V617Fcases[2] n=25	Confirmed PV cases [2] n=33	Controls n=473
Ate grilled or BBQ beef, lamb, or pork all year long					
No	29 (64%)	24 (69%)	15 (65%)	22 (71%)	304 (72%)
Yes	16 (36%)	11 (31%)	8 (35%)	9 (29%)	117 (28%)
Missing	10	6	2	2	52
Ever ate grilled, or BBQ poultry or fish					
Yes	36 (65%)	26 (63%)	16 (64%)	21 (64%)	407 (86%)
No	19 (35%)	15 (37%)	9 (36%)	12 (36%)	66 (14%)
Ate grilled, or BBQ poultry or fish in the winter					
No	20 (56%)	15 (58%)	7 (44%)	12 (57%)	285 (70%)
Yes	16 (44%)	11 (42%)	9 (56%)	9 (43%)	122 (30%)
Missing	19	15	9	12	66
Ate grilled, or BBQ poultry or fish in the spring					
No	19 (53%)	14 (54%)	7 (44%)	11 (52%)	215 (53%)
Yes	17 (47%)	12 (46%)	9 (56%)	10 (48%)	192 (47%)
Missing	19	15	9	12	66
Ate grilled, or BBQ poultry or fish in the summer					
No	0	0	0	0	3 (1%)
Yes	36 (100%)	26 (100%)	16 (100%)	21 (100%)	404 (99%)
Missing	19	15	9	12	66

Table 2.5. Cooked meat diet history[1] (cont'd)

	All cases[2] n=55	Confirmed cases[2] n=41	Confirmed JAK2 V617Fcases[2] n=25	Confirmed PV cases[2] n=33	Controls n=473
Ate grilled, or BBQ poultry or fish in the fall					
No	18 (50%)	13 (50%)	6 (38%)	10 (48%)	229 (56%)
Yes	18 (50%)	13 (50%)	10 (63%)	11 (52%)	178 (44%)
Missing	19	15	9	12	66
Ate grilled, or BBQ poultry or fish all year long					
No	20 (56%)	15 (58%)	7 (44%)	12 (57%)	287 (71%)
Yes	16 (44%)	11 (42%)	9 (56%)	9 (43%)	120 (29%)
Missing	19	15	9	12	66
Ever ate smoked beef, lamb, or pork					
Yes	42 (76%)	31 (76%)	19 (76%)	27 (82%)	436 (92%)
No	13 (24%)	10 (24%)	6 (24%)	6 (18%)	37 (8%)
Ate smoked beef, lamb, or pork in the winter					
No	1 (2%)	1 (3%)	0	1 (4%)	36 (8%)
Yes	41 (98%)	30 (97%)	19 (100%)	26 (96%)	400 (92%)
Missing	13	10	6	6	37
Ate smoked beef, lamb, or pork in the spring					
No	3 (7%)	3 (10%)	1 (5%)	2 (7%)	39 (9%)
Yes	39 (93%)	28 (90%)	18 (95%)	25 (93%)	397 (91%)
Missing	13	10	6	6	37

Table 2.5. Cooked meat diet history[1] (cont'd)

	All cases [2] n=55	Confirmed cases [2] n=41	Confirmed JAK2 V617Fcases[2] n=25	Confirmed PV cases [2] n=33	Controls n=473
Ate smoked beef, lamb, or pork in the summer					
No	5 (12%)	5 (16%)	2 (11%)	4 (15%)	31 (7%)
Yes	37 (88%)	26 (84%)	17 (89%)	23 (85%)	405 (93%)
Missing	13	10	6	6	37
Ate smoked beef, lamb, or pork in the fall					
No	3 (7%)	3 (10%)	1 (5%)	2 (7%)	43 (10%)
Yes	39 (93%)	28 (90%)	18 (95%)	25 (93%)	393 (90%)
Missing	13	10	6	6	37
Ate smoked beef, lamb, or pork all year long					
No	5 (12%)	5 (16%)	2 (11%)	4 (15%)	58 (13%)
Yes	37 (88%)	26 (84%)	17 (89%)	23 (85%)	378 (87%)
Missing	13	10	6	6	37
Ever ate smoked poultry or fish					
Yes	15 (28%)	12 (30%)	8 (33%)	11 (34%)	140 (30%)
No	39 (72%)	28 (70%)	16 (67%)	21 (66%)	332 (70%)
Don't know	1	1	1	1	1
Ate smoked poultry or fish in the winter					
No	1 (7%)	1 (8%)	0	1 (9%)	28 (20%)
Yes	14 (93%)	11 (92%)	8 (100%)	10 (91%)	112 (80%)
Missing	40	29	17	22	333

Table 2.5. Cooked meat diet history[1] (cont'd)

	All cases [2] n=55	Confirmed cases [2] n=41	Confirmed JAK2 V617Fcases[2] n=25	Confirmed PV cases [2] n=33	Controls n=473
Ate smoked poultry or fish in the spring					
No	3 (20%)	3 (25%)	1 (13%)	2 (18%)	35 (25%)
Yes	12 (80%)	9 (75%)	7 (88%)	9 (82%)	105 (75%)
Missing	40	29	17	22	333
Ate smoked poultry or fish in the summer					
No	3 (20%)	3 (25%)	1 (13%)	2 (18%)	29 (21%)
Yes	12 (80%)	9 (75%)	7 (88%)	9 (82%)	111 (79%)
Missing	40	29	17	22	333
Ate smoked poultry or fish in the fall					
No	2 (13%)	2 (17%)	1 (13%)	1 (9%)	18 (13%)
Yes	13 (87%)	10 (83%)	7 (88%)	10 (91%)	122 (87%)
Missing	40	29	17	22	333
Ate smoked poultry or fish all year					
No	3 (20%)	3 (25%)	1 (13%)	2 (18%)	40 (29%)
Yes	12 (80%)	9 (75%)	7 (88%)	9 (82%)	100 (71%)
Missing	40	29	17	22	333

1. Assessed through questions estimating grilled and barbequed meat intake during different time periods, number and %
2. All cases include PV, ET and PMF, confirmed cases by expert panel include PV, ET, and PMF, confirmed JAK2 V617F mutation with PV, ET or PMF by expert panel, confirmed by expert panel PV only

Table 2.6 Cooked meat diet history logistic regression models [1]

	All cases OR [2] n=55	All cases aOR [3] n=55	Confirmed cases OR [2] n=41	Confirmed cases aOR [3] n=41	Confirmed JAK2 V617F OR [2] n=25	Confirmed JAK2 V617F aOR [3] n=25	Confirmed PV OR [2] n=33	Confirmed PV aOR [3] n=33
Ever ate grilled, BBQ, or smoked meat or fish								
Yes	0.5 (0.1, 1.7)	0.5 (0.1, 2.0)	0.3 (0.1, 1.2)	0.4 (0.1, 1.5)	--	--	0.4 (0.1, 1.9)	0.5 (0.1, 2.6)
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Ate grilled, BBQ, or smoked meat or fish in the winter								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.2 (0.6, 2.2)	1.4 (0.7, 2.6)	1.2 (0.6, 2.3)	1.3 (0.6, 2.6)	1.1 (0.4, 2.5)	1.2 (0.5, 3.0)	0.9 (0.4, 2.0)	0.9 (0.4, 2.2)
Ate grilled, BBQ, smoked meat or fish in the spring								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	0.7 (0.4, 1.3)	0.9 (0.5, 1.6)	0.8 (0.4, 1.5)	0.9 (0.4, 1.8)	0.9 (0.4, 2.1)	1.2 (0.5, 2.7)	0.7 (0.3, 1.4)	0.7 (0.3, 1.6)

Table 2.6 Cooked meat diet history logistic regression models [1] (cont'd)

	All cases OR [2] n=55	All cases aOR [3] n=55	Confirmed cases OR [2] n=41	Confirmed cases aOR [3] n=41	Confirmed JAK2 V617F OR [2] n=25	Confirmed JAK2 V617F aOR [3] n=25	Confirmed PV OR [2] n=33	Confirmed PV aOR [3] n=33
Ate grilled, BBQ, smoked meat or fish in the fall								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	0.8 (0.4, 1.5)	0.9 (0.5, 1.8)	0.8 (0.4, 1.6)	0.9 (0.4, 1.9)	0.9 (0.4, 2.1)	1.2 (0.5, 2.7)	0.7 (0.3, 1.4)	0.7 (0.3, 1.6)
Ate grilled, BBQ, smoked meat or fish all year								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.1 (0.6, 2.1)	1.3 (0.7, 2.4)	1.1 (0.5, 2.2)	1.2 (0.6, 2.4)	1.1 (0.5, 2.6)	1.3 (0.5, 3.1)	0.8 (0.4, 1.9)	0.8 (0.3, 1.9)
Ever ate grilled or BBQ beef, lamb, or pork								
Yes	0.6 (0.3, 1.2)	0.7 (0.3, 1.5)	0.7 (0.3, 1.8)	0.9 (0.4, 2.3)	1.4 (0.3, 6.2)	1.9 (0.4, 8.3)	1.9 (0.4, 8.2)	2.6 (0.6, 11.7)
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Ate grilled or BBQ beef, lamb, or pork in the winter								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.4 (0.7, 2.6)	1.4 (0.7, 2.7)	1.1 (0.5, 2.4)	1.1 (0.5, 2.4)	1.3 (0.5, 3.2)	1.4 (0.5, 3.6)	1.0 (0.5, 2.3)	0.9 (0.4, 2.1)

Table 2.6 Cooked meat diet history logistic regression models [1] (cont'd)

	All cases OR [2] n=55	All cases aOR [3] n=55	Confirmed cases OR [2] n=41	Confirmed cases aOR [3] n=41	Confirmed JAK2 V617F OR [2] n=25	Confirmed JAK2 V617F aOR [3] n=25	Confirmed PV OR [2] n=33	Confirmed PV aOR [3] n=33
Ate grilled or BBQ beef, lamb, or pork in the spring								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.0 (0.6, 1.9)	1.1 (0.6, 2.2)	0.9 (0.4, 1.8)	0.9 (0.4, 1.9)	1.1 (0.5, 2.5)	1.3 (0.5, 3.3)	0.9 (0.4, 1.8)	0.8 (0.4, 1.8)
Ate grilled or BBQ beef, lamb, or pork in the summer								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	--	--	--	--	--	--	--	--
Ate grilled or BBQ beef, lamb, or pork in the fall								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.0 (0.5, 1.9)	1.0 (0.5, 2.0)	0.9 (0.5, 1.8)	0.9 (0.4, 1.9)	1.1 (0.5, 2.4)	1.2 (0.5, 3.0)	0.9 (0.4, 1.8)	0.8 (0.4, 1.8)
Ate grilled or BBQ beef, lamb, or pork all year long								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.4 (0.8, 2.7)	1.4 (0.7, 2.9)	1.2 (0.6, 2.5)	1.1 (0.5, 2.5)	1.4 (0.6, 3.4)	1.5 (0.6, 3.9)	1.1 (0.5, 2.4)	0.9 (0.4, 2.2)

Table 2.6 Cooked meat diet history logistic regression models [1] (cont'd)

	All cases OR [2] n=55	All cases aOR [3] n=55	Confirmed cases OR [2] n=41	Confirmed cases aOR [3] n=41	Confirmed JAK2 V617F OR [2] n=25	Confirmed JAK2 V617F aOR [3] n=25	Confirmed PV OR [2] n=33	Confirmed PV aOR [3] n=33
Ever ate grilled, or BBQ poultry or fish								
Yes	0.3 (0.2, 0.6)	0.4 (0.2, 0.7)	0.3 (0.1, 0.6)	0.3 (0.2, 0.7)	0.3 (0.1, 0.7)	0.4 (0.1, 0.9)	0.3 (0.1, 0.6)	0.4 (0.2, 0.8)
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Ate grilled, or BBQ poultry or fish in the winter								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.9 (0.9, 3.7)	2.0 (1.0, 4.1)	1.7 (0.8, 3.8)	1.7 (0.7, 4.0)	3.0 (1.1, 8.2)	3.5 (1.2, 10.0)	1.8 (0.7, 4.3)	1.6 (0.6, 4.1)
Ate grilled, or BBQ poultry or fish in the spring								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.0 (0.5, 2.0)	1.1 (0.5, 2.3)	1.0 (0.4, 2.1)	1.0 (0.4, 2.3)	1.4 (0.5, 3.9)	1.8 (0.6, 5.2)	1.0 (0.4, 2.4)	1.0 (0.4, 2.5)
Ate grilled, or BBQ poultry or fish in the summer								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	--	--	--	--	--	--	--	--

Table 2.6 Cooked meat diet history logistic regression models [1] (cont'd)

	All cases OR [2] n=55	All cases aOR [3] n=55	Confirmed cases OR [2] n=41	Confirmed cases aOR [3] n=41	Confirmed JAK2 V617F OR [2] n=25	Confirmed JAK2 V617F aOR [3] n=25	Confirmed PV OR [2] n=33	Confirmed PV aOR [3] n=33
Ate grilled, or BBQ poultry or fish in the fall								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.3 (0.7, 2.5)	1.4 (0.7, 2.9)	1.3 (0.6, 2.8)	1.3 (0.6, 3.1)	2.1 (0.8, 6.0)	2.6 (0.9, 7.6)	1.4 (0.6, 3.4)	1.4 (0.6, 3.5)
Ate grilled, or BBQ poultry or fish all year long								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.9 (1.0, 3.8)	2.1 (1.0, 4.2)	1.8 (0.8, 3.9)	1.8 (0.8, 4.1)	3.1 (1.1, 8.4)	3.6 (1.2, 10.4)	1.8 (0.7, 4.4)	1.7 (0.7, 4.3)
Ever ate smoked beef, lamb, or pork								
Yes	0.3 (0.1, 0.6)	0.3 (0.1, 0.5)	0.3 (0.1, 0.6)	0.2 (0.1, 0.6)	0.3 (0.1, 0.7)	0.2 (0.1, 0.7)	0.4 (0.1, 1.0)	0.4 (0.1, 1.0)
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Ate smoked beef, lamb, or pork in the winter								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	3.7 (0.5, 27.6)	4.2 (0.6, 32.1)	2.7 (0.4, 20.4)	3.1 (0.4, 24.2)	--	--	2.3 (0.3, 17.7)	2.7 (0.3, 20.9)

Table 2.6 Cooked meat diet history logistic regression models [1] (cont'd)

	All cases OR [2] n=55	All cases aOR [3] n=55	Confirmed cases OR [2] n=41	Confirmed cases aOR [3] n=41	Confirmed JAK2 V617F OR [2] n=25	Confirmed JAK2 V617F aOR [3] n=25	Confirmed PV OR [2] n=33	Confirmed PV aOR [3] n=33
Ate smoked beef, lamb, or pork in the spring								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.3 (0.4, 4.3)	1.4 (0.4, 5.0)	0.9 (0.3, 3.2)	1.0 (0.3, 3.6)	1.8 (0.2, 13.6)	2.2 (0.3, 17.6)	1.2 (0.3, 5.4)	1.4 (0.3, 6.1)
Ate smoked beef, lamb, or pork in the summer								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	0.6 (0.2, 1.5)	0.6 (0.2, 1.8)	0.4 (0.1, 1.1)	0.4 (0.2, 1.3)	0.7 (0.1, 2.9)	0.9 (0.2, 4.0)	0.4 (0.1, 1.4)	0.5 (0.1, 1.5)
Ate smoked beef, lamb, or pork in the fall								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.4 (0.4, 4.8)	1.4 (0.4, 4.8)	1.0 (0.3, 3.5)	1.0 (0.3, 3.5)	2.0 (0.3, 15.1)	2.1 (0.3, 16.4)	1.4 (0.3, 6.0)	1.2 (0.3, 5.6)
Ate smoked beef, lamb, or pork all year long								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.1 (0.4, 3.0)	1.2 (0.5, 3.3)	0.8 (0.3, 2.2)	0.9 (0.3, 2.4)	1.3 (0.3, 5.8)	1.6 (0.3, 7.2)	0.9 (0.3, 2.6)	0.9 (0.3, 2.8)

Table 2.6 Cooked meat diet history logistic regression models [1] (cont'd)

	All cases OR [2] n=55	All cases aOR [3] n=55	Confirmed cases OR [2] n=41	Confirmed cases aOR [3] n=41	Confirmed JAK2 V617F OR [2] n=25	Confirmed JAK2 V617F aOR [3] n=25	Confirmed PV OR [2] n=33	Confirmed PV aOR [3] n=33
Ever ate smoked poultry or fish								
Yes	0.9 (0.5, 1.7)	1.0 (0.5, 1.9)	1.0 (0.5, 2.1)	1.1 (0.5, 2.3)	1.2 (0.5, 2.8)	1.3 (0.5, 3.3)	1.2 (0.6, 2.6)	1.4 (0.6, 3.0)
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Ate smoked poultry or fish in the winter								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	3.5 (0.4, 27.8)	3.3 (0.4, 28.7)	2.7 (0.3, 22.2)	2.5 (0.3, 22.3)	--	--	2.5 (0.3, 20.3)	2.5 (0.3, 22.2)
Ate smoked poultry or fish in the spring								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.3 (0.4, 5.0)	1.2 (0.3, 4.8)	1.0 (0.3, 3.9)	0.9 (0.2, 3.7)	2.3 (0.3, 19.6)	2.1 (0.2, 19.7)	1.5 (0.3, 7.3)	1.3 (0.3, 7.1)
Ate smoked poultry or fish in the summer								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.0 (0.3, 3.9)	0.8 (0.2, 3.3)	0.8 (0.2, 3.1)	0.6 (0.1, 2.6)	1.8 (0.2, 15.5)	1.4 (0.2, 13.2)	1.2 (0.2, 5.7)	0.9 (0.2, 4.9)

Table 2.6 Cooked meat diet history logistic regression models [1] (cont'd)

	All cases OR [2] n=55	All cases aOR [3] n=55	Confirmed cases OR [2] n=41	Confirmed cases aOR [3] n=41	Confirmed JAK2 V617F OR [2] n=25	Confirmed JAK2 V617F aOR [3] n=25	Confirmed PV OR [2] n=33	Confirmed PV aOR [3] n=33
Ate smoked poultry or fish in the fall								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.0 (0.2, 4.6)	0.9 (0.2, 5.2)	0.7 (0.1, 3.6)	0.7 (0.1, 4.1)	1.0 (0.1, 8.9)	1.0 (0.1, 9.9)	1.5 (0.2, 12.2)	1.4 (0.2, 13.5)
Ate smoked poultry or fish all year								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.6 (0.4, 6.0)	1.3 (0.3, 5.3)	1.2 (0.3, 4.7)	1.0 (0.2, 4.1)	2.8 (0.3, 23.5)	2.3 (0.3, 21.3)	1.8 (0.4, 8.7)	1.5 (0.3, 8.0)

1. Assessed through questions estimating grilled and barbequed meat intake during different time periods,
2. All cases include PV, ET and PMF, confirmed cases by expert panel include PV, ET, and PMF, confirmed JAK2 V617F mutation with PV, ET or PMF by expert panel, confirmed by expert panel PV only
3. Adjusted (for age, sex and county) ORs (95% confidence intervals)

Table 2.7 Residential history [1]

	All cases [2] n=55	Confirmed cases [2] n=41	Confirmed JAK2 V617Fcases [2] n=25	Confirmed PV cases [2] n=33	Controls n=473
Lived within 0.5 mile of a dump/landfill					
No	47 (85%)	37 (90%)	23 (92%	29 (88%)	377 (80%)
Yes	8 (15%)	4 (10%)	2 (8%)	4 (12%)	96 (20%)
Lived within 0.5 mile of a hazardous waste site					
No	51 (93%)	40 (98%)	25 (100%)	32 (97%)	436 (92%)
Yes	4 (7%)	1 (2%)	0	1 (3%)	37 (8%)
Lived within 0.5 mile of an airport					
No	48 (87%)	37 (90%)	24 (96%)	30 (91%)	396 (84%)
Yes	7 (13%)	4 (10%)	1 (4%)	3 (9%)	77 (16%)
Lived within 0.5 mile of a farm					
No	30 (55%)	20 (49%)	13 (52%)	16 (48%)	262 (55%)
Yes	25 (45%)	21 (51%)	12 (48%)	17 (52%)	211 (45%)
Lived within 0.5 mile of a nursery/greenhouse					
No	42 (76%)	33 (80%)	20 (80%)	26 (79%)	343 (73%)
Yes	13 (24%)	8 (20%)	5 (20%)	7 (21%)	130 (27%)

Table 2.7 Residential history [1] (cont'd)

	All cases [2] n=55	Confirmed cases [2] n=41	Confirmed JAK2 V617Fcases [2] n=25	Confirmed PV cases [2] n=33	Controls n=473
Lived within 0.5 mile of a golf course					
No	49 (89%)	36 (88%)	22 (88%)	28 (85%)	390 (82%)
Yes	6 (11%)	5 (12%)	3 (12%)	5 (15%)	83 (18%)
Lived within 0.5 mile of a railroad track					
No	24 (44%)	19 (46%)	11 (44%)	15 (45%)	163 (34%)
Yes	31 (56%)	22 (54%)	14 (56%)	18 (55%)	310 (66%)
Lived within 0.5 mile of a gas station					
No	18 (33%)	16 (39%)	9 (36%)	13 (39%)	84 (18%)
Yes	37 (67%)	25 (61%)	16 (64%)	20 (61%)	389 (82%)
Lived within 0.5 mile of a high voltage tower					
No	43 (78%)	34 (83%)	22 (88%)	26 (79%)	369 (78%)
Yes	12 (22%)	7 (17%)	3 (12%)	7 (21%)	104 (22%)

Table 2.7 Residential history [1] (cont'd)

	All cases [2] n=55	Confirmed cases [2] n=41	Confirmed JAK2 V617Fcases [2] n=25	Confirmed PV cases [2] n=33	Controls n=473
Lived within 0.5 mile of a incinerator					
No	55 (100%)	41 (100%)	25 (100%)	33 (100%)	450 (95%)
Yes	0	0	0	0	23 (5%)
Lived within 0.5 mile of a factory/industrial plant					
No	33 (60%)	26 (63%)	14 (56%)	18 (55%)	219 (46%)
Yes	22 (40%)	15 (37%)	11 (44%)	15 (45%)	254 (54%)
Lived within 0.5 mile of a quarry or mine					
No	31 (56%)	23 (56%)	13 (52%)	17 (52%)	293 (62%)
Yes	24 (44%)	18 (44%)	12 (48%)	16 (48%)	180 (38%)
Lived within 0.5 mile of a coal power plant					
No	54 (98%)	40 (98%)	24 (96%)	32 (97%)	452 (96%)
Yes	1 (2%)	1 (2%)	1 (4%)	1 (3%)	21 (4%)

Table 2.7 Residential history [1] (cont'd)

	All cases [2] n=55	Confirmed cases [2] n=41	Confirmed JAK2 V617Fcases [2] n=25	Confirmed PV cases [2] n=33	Controls n=473
Lived within 0.5 mile of a nuclear power plant					
No	55 (100%)	41 (100%)	25 (100%)	33 (100%)	469 (99%)
Yes	0	0	0	0	4 (1%)

1 Every primary residence that they have lived in for 6 months or longer beginning with current residence and work backwards, listing all residences back to age 21, number and %

2. All cases include PV, ET and PMF, confirmed cases by expert panel include PV, ET, and PMF, confirmed JAK2 V617F mutation with PV, ET or PMF by expert panel, confirmed by expert panel PV only

Table 2.8 Residential history logistic regression models [1]

	All cases OR [2] (95%CI) n=55	All cases aOR [3] (95%CI) n=55	Confirmed cases OR [2] (95%CI) n=41	Confirmed cases aOR [3] (95%CI) n=41	Confirmed JAK2 V617F OR [2] (95%CI) n=25	Confirmed JAK2 V617F aOR [3] (95%CI) n=25	Confirmed PV OR [2] (95%CI) n=33	Confirmed PV aOR [3] (95%CI) n=33
Lived within 0.5 mile of a dump/landfill								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	0.7 (0.3, 1.5)	0.6 (0.3, 1.4)	0.4 (0.1, 1.2)	0.4 (0.1, 1.1)	0.3 (0.1, 1.5)	0.3 (0.1, 1.4)	0.5 (0.2, 1.6)	0.5 (0.2, 1.5)
Lived within 0.5 mile of a hazardous waste site								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	0.9 (0.3, 2.7)	1.0 (0.3, 3.0)	0.3 (0.0, 2.2)	0.3 (0.0, 2.4)	--	--	0.4 (0.0, 2.8)	0.4 (0.0, 2.8)
Lived within 0.5 mile of an airport								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	0.8 (0.3, 1.7)	0.8 (0.4, 2.0)	0.6 (0.2, 1.6)	0.6 (0.2, 1.8)	0.2 (0.0, 1.6)	0.2 (0.0, 1.8)	0.5 (0.2, 1.7)	0.6 (0.2, 2.1)
Lived within 0.5 mile of a farm								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.0 (0.6, 1.8)	1.0 (0.6, 1.9)	1.3 (0.7, 2.5)	1.3 (0.7, 2.5)	1.1 (0.5, 2.6)	1.2 (0.5, 2.8)	1.3 (0.7, 2.7)	1.3 (0.6, 2.6)
Lived within 0.5 mile of a nursery/greenhouse								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	0.8 (0.4, 1.6)	0.9 (0.5, 1.7)	0.6 (0.3, 1.4)	0.7 (0.3, 1.6)	0.7 (0.2, 1.8)	0.7 (0.3, 2.0)	0.7 (0.3, 1.7)	0.8 (0.3, 1.9)

Table 2.8 Residential history logistic regression models [1] (cont'd)

	All cases OR [2] (95%CI) n=55	All cases aOR [3] (95%CI) n=55	Confirmed cases OR [2] (95%CI) n=41	Confirmed cases aOR [3] (95%CI) n=41	Confirmed JAK2 V617F OR [2] (95%CI) n=25	Confirmed JAK2 V617F aOR [3] (95%CI) n=25	Confirmed PV OR [2] (95%CI) n=33	Confirmed PV aOR [3] (95%CI) n=33
Lived within 0.5 mile of a golf course								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	0.6 (0.2, 1.4)	0.6 (0.3, 1.5)	0.7 (0.2, 1.7)	0.7 (0.3, 1.9)	0.6 (0.2, 2.2)	0.7 (0.2, 2.4)	0.8 (0.3, 2.2)	1.0 (0.4, 2.7)
Lived within 0.5 mile of a railroad track								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	0.7 (0.4, 1.2)	0.7 (0.4, 1.2)	0.6 (0.3, 1.2)	0.6 (0.3, 1.1)	0.7 (0.3, 1.5)	0.6 (0.3, 1.5)	0.6 (0.3, 1.3)	0.6 (0.3, 1.3)
Lived within 0.5 mile of a gas station								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	0.4 (0.2, 0.8)	0.5 (0.2, 0.8)	0.3 (0.2, 0.7)	0.3 (0.2, 0.7)	0.4 (0.2, 0.9)	0.4 (0.2, 1.0)	0.3 (0.2, 0.7)	0.3 (0.2, 0.7)
Lived within 0.5 mile of a high voltage tower								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.0 (0.5, 1.9)	1.0 (0.5, 2.1)	0.7 (0.3, 1.7)	0.8 (0.3, 1.8)	0.5 (0.1, 1.6)	0.5 (0.1, 1.8)	1.0 (0.4, 2.3)	1.0 (0.4, 2.4)
Lived within 0.5 mile of a incinerator								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	--	--	--	--	--	--	--	--

Table 2.8 Residential history logistic regression models [1] (cont'd)

	All cases OR [2] (95%CI) n=55	All cases aOR [3] (95%CI) n=55	Confirmed cases OR [2] (95%CI) n=41	Confirmed cases aOR [3] (95%CI) n=41	Confirmed JAK2 V617F OR [2] (95%CI) n=25	Confirmed JAK2 V617F aOR [3] (95%CI) n=25	Confirmed PV OR [2] (95%CI) n=33	Confirmed PV aOR [3] (95%CI) n=33
Lived within 0.5 mile of a factory/industrial plant								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	0.6 (0.3, 1.0)	0.6 (0.3, 1.0)	0.5 (0.3, 1.0)	0.5 (0.3, 1.0)	0.7 (0.3, 1.5)	0.7 (0.3, 1.6)	0.7 (0.4, 1.5)	0.7 (0.3, 1.5)
Lived within 0.5 mile of a quarry or mine								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	1.3 (0.7, 2.2)	1.3 (0.7, 2.4)	1.3 (0.7, 2.4)	1.3 (0.7, 2.6)	1.5 (0.7, 3.4)	1.5 (0.7, 3.5)	1.5 (0.8, 3.1)	1.7 (0.8, 3.4)
Lived within 0.5 mile of a coal power plant								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	0.4 (0.1, 3.0)	0.3 (0.0, 2.2)	0.5 (0.1, 4.1)	0.4 (0.0, 3.1)	0.9 (0.1, 7.0)	0.7 (0.1, 5.3)	0.7 (0.1, 5.2)	0.4 (0.1, 3.3)

1. Every primary residence that they have lived in for 6 months or longer beginning with current residence and work backwards, listing all residences back to age 21
2. All cases include PV, ET and PMF, confirmed cases by expert panel include PV, ET, and PMF, confirmed JAK2 V617F mutation with PV, ET or PMF by expert panel, confirmed by expert panel PV only
3. Adjusted (for age, sex and county) Odds Ratio (95% confidence intervals)

Table 2.9. Other lifestyle behaviors [1]

	All cases [2] n=55	Confirmed cases [2] n=41	Confirmed JAK2 V617Fcases [2] n=25	Confirmed PV cases [2] n=33	Controls n=473
Ever use hair color					
Yes	26 (47%)	20 (49%)	14 (56%)	15 (45%)	267 (56%)
No	29 (53%)	21 (51%)	11 (44%)	18 (55%)	206 (44%)
Hair coloring frequency					
Weekly (or more)	1 (4%)	1 (5%)	0	1 (7%)	3 (1%)
Monthly	11 (42%)	8 (40%)	5 (36%)	6 (40%)	128 (48%)
Yearly (A few times a year or less)	14 (54%)	11 (55%)	9 (64%)	8 (53%)	134 (51%)
Don't know	0	0	0	0	2
Ever visited the Still Creek Reservoir					
Yes	4 (7%)	2 (5%)	1 (4%)	2 (6%)	30 (6%)
No	51 (93%)	39 (95%)	24 (96%)	31 (94%)	438 (94%)
Don't know	0	0	0	0	5
Ever take soil from Still Creek Reservoir					
Yes	0	0	0	0	1 (3%)
No	4 (100%)	2 (100%)	1 (100%)	2 (100%)	29 (97%)
Ever eaten fish from other local creeks or lakes					
Yes	16 (29%)	13 (32%)	7 (28%)	10 (30%)	181 (39%)
No	39 (71%)	28 (68%)	18 (72%)	23 (70%)	286 (61%)
Don't know	0	0	0	0	6

1. Other behaviors, including hair coloring, visiting Still Creek Reservoir, and eating fish from other local lakes or creeks, number and % with 95% confidence intervals
2. All cases include PV, ET and PMF, confirmed cases by expert panel include PV, ET, and PMF, confirmed JAK2 V617F mutation with PV, ET or PMF by expert panel, confirmed by expert panel PV only

Table 2.10 Other lifestyle behavior logistic regression models [1]

	All cases OR [2] (95%CI)	All cases aOR [3] (95%CI)	Confirmed cases OR [2] (95%CI)	Confirmed cases aOR [3] (95%CI)	Confirmed JAK2 V617F OR [2] (95%CI)	Confirmed JAK2 V617F aOR [3] (95%CI)	Confirmed PV OR [2] (95%CI)	Confirmed PV aOR [3] (95%CI)
Ever use hair color								
Yes	0.7 (0.4, 1.2)	1.0 (0.5, 2.2)	0.7 (0.4, 1.4)	1.1 (0.5, 2.8)	1.0 (0.4, 2.2)	1.9 (0.6, 6.0)	0.6 (0.3, 1.3)	1.3 (0.5, 3.5)
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Ever visited the Still Creek Reservoir								
Yes	1.1 (0.4, 3.4)	1.0 (0.3, 3.2)	0.7 (0.2, 3.3)	0.6 (0.1, 2.9)	0.6 (0.1, 4.7)	0.5 (0.1, 4.5)	0.9 (0.2, 4.1)	0.7 (0.1, 3.1)
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Ever eaten fish from other local creeks or lakes								
Yes	0.6 (0.4, 1.2)	0.7 (0.4, 1.3)	0.7 (0.4, 1.5)	0.8 (0.4, 1.5)	0.6 (0.3, 1.5)	0.7 (0.3, 1.7)	0.7 (0.3, 1.5)	0.7 (0.3, 1.6)
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Lived within 0.5 mile of a coal power plant								
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Yes	0.4 (0.1, 3.0)	0.3 (0.0, 2.2)	0.5 (0.1, 4.1)	0.4 (0.0, 3.1)	0.9 (0.1, 7.0)	0.7 (0.1, 5.3)	0.7 (0.1, 5.2)	0.4 (0.1, 3.3)

1. Other behaviors, including hair coloring, visiting Still Creek Reservoir, and eating fish from other local lakes or creeks
2. All cases include PV, ET and PMF, confirmed cases by expert panel include PV, ET, and PMF, confirmed JAK2 V617F mutation with PV, ET or PMF by expert panel, confirmed by expert panel PV only
3. Adjusted (for age, sex and county) Odds Ratio (95% confidence intervals)

Table 2.11 Smoking history [1] (cont'd)

Smoking history	All cases [2] n=55	Confirmed cases [2] n=41	Confirmed JAK2 V617Fcases [2] n=25	Confirmed PV cases [2] n=33	Controls n=473
Ever smoked					
Yes	27 (49%)	18 (44%)	8 (32%)	15 (45%)	237 (50%)
No	28 (51%)	23 (56%)	17 (68%)	18 (55%)	236 (50%)
Ever smoked cigarettes					
Yes	27 (100%)	18 (100%)	8 (100%)	15 (100%)	234 (99%)
No	0	0	0	0	3 (1%)
Cigarette pack-years					
Non smoker	28 (51%)	23 (56%)	17 (68%)	18 (55%)	239 (51%)
Light smoker (<15 py)	13 (24%)	10 (24%)	6 (24%)	8 (24%)	115 (24%)
Heavy smoker (>=15 py)	14 (25%)	8 (20%)	2 (8%)	7 (21%)	119 (25%)
Cigarette pack-years (among smokers)					
Light smoker (<15 py)	13 (48%)	10 (56%)	6 (75%)	8 (53%)	115 (49%)
Heavy smoker (>=15 py)	14 (52%)	8 (44%)	2 (25%)	7 (47%)	119 (51%)
Smoked filtered or nonfiltered cigarettes					
Filtered	21 (78%)	13 (72%)	6 (75%)	10 (67%)	203 (87%)
Non-filtered	6 (22%)	5 (28%)	2 (25%)	5 (33%)	30 (13%)
Both	0	0	0	0	1 (0%)
Smoked manufactured or hand rolled cigarettes					
Manufactured (store bought)	22 (81%)	16 (89%)	7 (88%)	13 (87%)	215 (92%)
Hand rolled	1 (4%)	0	0	0	2 (1%)
Both	4 (15%)	2 (11%)	1 (13%)	2 (13%)	17 (7%)

Table 2.11 Smoking history [1] (cont'd)

Smoking history	All cases [2] n=55	Confirmed cases [2] n=41	Confirmed JAK2 V617Fcases [2] n=25	Confirmed PV cases [2] n=33	Controls n=473
Number of smoking episodes					
1	27 (100%)	18 (100%)	8 (100%)	15 (100%)	202 (86%)
2	0	0	0	0	24 (10%)
3	0	0	0	0	7 (3%)
4	0	0	0	0	1 (0%)
5	0	0	0	0	0
Number of cessation attempts					
0	6 (22%)	4 (22%)	1 (13%)	3 (20%)	51 (22%)
1	21 (78%)	14 (78%)	7 (88%)	12 (80%)	161 (69%)

1 .Detailed self-reported smoking history, number and %
2. All cases include PV, ET and PMF, confirmed cases by expert panel include PV, ET, and PMF, confirmed JAK2 V617F mutation with PV, ET or PMF by expert panel, confirmed by expert panel PV only

Table 2.12 Smoking history logistic regression models [1]

	All cases OR [2] (95%CI) n=55	All cases aOR [3] (95%CI) n=55	Confirmed cases OR [2] (95%CI) n=41	Confirmed cases aOR [3] (95%CI) n=41	Confirmed JAK2 V617F OR [2] (95%CI) n=25	Confirmed JAK2 V617F aOR [3] (95%CI) n=25	Confirmed PV OR [2] (95%CI) n=33	Confirmed PV aOR [3] (95%CI) n=33
Ever smoked								
Yes	1.0 (0.5, 1.7)	0.8 (0.5, 1.5)	0.8 (0.4, 1.5)	0.7 (0.3, 1.3)	0.5 (0.2, 1.1)	0.4 (0.2, 1.0)	0.8 (0.4, 1.7)	0.7 (0.3, 1.5)
No	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Cigarette pack-years								
Non smoker	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Light smoker (<15 py)	1.0 (0.5, 1.9)	0.9 (0.5, 1.9)	0.9 (0.4, 2.0)	0.9 (0.4, 2.0)	0.7 (0.3, 1.9)	0.7 (0.3, 2.0)	0.9 (0.4, 2.2)	0.9 (0.4, 2.3)
Heavy smoker (>=15 py)	1.0 (0.5, 2.0)	0.8 (0.4, 1.7)	0.7 (0.3, 1.6)	0.6 (0.2, 1.3)	0.2 (0.1, 1.0)	0.2 (0.0, 0.8)	0.8 (0.3, 1.9)	0.6 (0.2, 1.5)
Cigarette pack-years (among smokers)								
Light smoker (<15 py)	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Heavy smoker (>=15 py)	1.0 (0.5, 2.3)	0.8 (0.4, 2.0)	0.8 (0.3, 2.0)	0.6 (0.2, 1.6)	0.3 (0.1, 1.6)	0.2 (0.0, 1.4)	0.8 (0.3, 2.4)	0.6 (0.2, 1.8)
Smoked filtered or nonfiltered cigarettes								
Filtered	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Non-filtered	1.9 (0.7, 5.2)	1.1 (0.4, 3.1)	2.6 (0.9, 7.8)	1.5 (0.5, 4.9)	2.3 (0.4, 11.7)	1.1 (0.2, 6.4)	3.4 (1.1, 10.6)	2.0 (0.6, 6.9)
Both	--	--	--	--	--	--	--	--

1.Detailed self-reported smoking history
2.All cases include PV, ET and PMF, confirmed cases by expert panel include PV, ET, and PMF, confirmed JAK2 V617F mutation with PV, ET or PMF by expert , confirmed by expert panel PV only 3. Adjusted (for age, sex and county) Odds Ratio (95% confidence intervals)

Chapter 3: The Role of Genotypes that Modify the Toxicity of Chemical Mutagens in the Risk for Myeloproliferative Neoplasms

Abstract

Introduction

The etiology of a rare category of myeloproliferative neoplasms (MPN)—bone marrow diseases with an excess of blood cells—is currently unknown. These MPNs (polycythemia vera (PV), essential thrombocythemia (ET) and primary myelofibrosis (PMF)) differ phenotypically but commonly share the same JAK2 V617F point mutation (JAK2 V617F), which is thought to be at least partly responsible for disease initiation or progression. Thus, the presence of JAK2 V617F mutation suspected to be pathognomonic of MPNs is included in this study as an outcome of interest.

Objectives

To investigate the potential for gene-environment interactions in aetiology of MPNs and JAK2 V617F using a biological pathway candidate gene approach.

Methods

We conducted a population-based case-control study among residents of three Pennsylvania counties where a cluster of PV was previously described. Subjects were included if they were born between 1921 and 1968 and resided in the three counties between 2000 and 2008 in 2011. Cases were identified from the Pennsylvania Cancer Registry and a previously completed cluster investigation we used multiple case categorizations, including all MPNs combined, JAK2 V617F cases, and PV cases only. Controls were selected based on eligibility screening following random digit dialling with the aim of obtaining an age and sex distribution that reflected that of the sources counties. A DNA sample was obtained from participants consenting to blood collection and genotyped for a panel of *a prior* selected environmentally sensitive genes. Blood samples were collected from 31 cases and 292 controls. Data were analyzed using logistic regression models that controlled for the design variables (age, sex, county) and evaluated one genetic variant at a time.

Results

Cases and controls were demographically similar. In analysis that examined the main effects of 14 environmentally sensitive genes, the presence of *NAT2* slow acetylator genotype, CYP1A2 and *GSTA1*, and *GSTM3* variants were associated with an increased risk for all MPNs combined (point estimates of adjusted (aORs) 2.7 to 4.6, with 95% C.I.s that excluded 1.0). Results were similar for analysis restricted to JAK2 V617F cases.

Conclusions

Our findings suggest that genotypes that modify the toxicity of aromatic and heterocyclic amines play a role in MPNs. Sources and types of exposures important to the pathway whereby *NAT2* or other genotypes modify risk of MPNs in this population remain unclear.

Introduction

Myeloproliferative neoplasms are rare cancers characterized by an overproduction of red blood cells and platelets. These includes polycythemia vera (PV), essential thrombocythemia (ET), and primary myelofibrosis (PMF) (Campbell & Green, 2005). PV patients expense an excess of red blood cells, >25% above predicted red cell mass (Campbell &Green, 2005). The overproduction of platelets in the bone marrow is seen in ET cases, and scarring of the bone occurs in PMF cases (the most severe of the three) (Tefferi et al., 2007). The World Health Organization classified MPNs and reporting on MPNs began in 2001. Before the WHO's revised diagnostic criterion in 2008, the National Cancer Institute estimated the incidence of PV at 2.8 per 100,000 persons per year for 2001-2004 (Kutti &Ridell, 2001). For ET the incidence rate was estimated at 1.5 cases per 100,000 per year (Kutti &Ridell, 2001). For PMF, only 0.4 cases per 100,000 persons per year were estimated (Kutti &Ridell, 2001). There are no known causes of MPNs (Anderson et al.,2012).

The JAK2 V617F mutation and its role in MPNs

The presence of the JAK2 V617F somatic point mutation is suspected to be pathognomonic of PV with clinical diagnosis of PV. JAK2 is a protein that acts as an on-and-off switch regulating bone marrow activity (Baxter, 2005). The acquired mutation is a single-base substitute that results in a valine to phenylalanine amino acid at position 617 on the JAK2 gene (Langabeer et al., 2007) and is seen in nearly all PV patients (>95%) and about half of those with ET or PMF, and less frequently in other hematologic diseases but not in any other cancers (Spivak et al., 2010; Kralovics et al., 2007; Tefferi et al., 2007). The JAK2 V617F mutation activates a tyrosine kinase complex in bone marrow normally responsible for regulating blood cell production through molecular signalling; the mutation disrupts the normal inhibition of growth, thus increasing blood

cell production (Spivak et al., 2010). The causes of the JAK2 V617F mutation are currently unknown (Kralovics et al., 2007).

Familial clustering of PV, ET, and PMF

There is strong evidence for an increased risk of developing an MPN in people who have a family member with an MPN (Anderson et al., 2012, Landgren et al., 2008). Familial clustering has been documented in a 24,577 first degree relatives of 11,039 MPN patients diagnosed from 1958 to 2005 in Sweden, which showed a 5 to 7-fold increase in risk of developing an MPN for first-degree relatives of MPN patients (Landgren et al., 2008). The role of inherited genetics has been suspected to influence both MPN phenotype and susceptibility, including potential germ-line mutations yet to be identified (Andrikovics et al., 2010).

Recently a germ-line haplotype, a sequence of single nucleotide polymorphisms (SNPs) on the loci designated as the 46/1 haplotype, has been identified as strongly associated with JAK2 V617F status (Jones et al., 2009; Kilpivaara et al., 2009) (Olcaydu et al., 2009). Three genome-wide studies have identified this germ-line 46/1 haplotype in European populations in the range of 5%-41% associated with positivity MPNs (Jones et al., 2009; Olcaydu et al., 2009; Kilpivaara et al., 2009). The 46/1 haplotype is considered to be the only known risk factor for the JAK2 V617F mutation in PV patients (Jones et al., 2009; Andrikovics et al., 2010).

Environmental exposures and MPNs

Since the discovery of the acquired and germ-line mutations in MPN patients, the notion that environmental exposures that can produce such mutations may play a role in the increased risk of developing an MPN became plausible. The evidence that exists in the literature to date suggests that some occupational exposures to may be associated with

MPNs. There were three cohort studies looking at PV or myelofibrosis that reported higher mortality rates. Zoloth reported in 1986 higher than expected standardized mortality rates (SMRs) among commercial pressmen n=2,500 in US, 120 cases per 100,000 compared to incidence in population of 2 per 100,000 with four deaths from PV (Zoloth et al., 1986). In a study of poultry workers in Baltimore, a proportional mortality rate of 4.9 (95%CI: 1.4,017.2) with myelofibrosis was reported for only two cases out of a cohort of 28,900 poultry workers from 1954 to 1979 with follow up until 2003 (Johnson et al., 2010). Kaplan et al., reported an elevated SMR of petroleum refiner workers of 455 (95%CI: 1.20, 11.64) (Kaplan et al., 1986). Although in these studies that detected statistically significant results, there were only 14 cases of MPNs, of which only six were PV, and none were ET or PMF.

Different occupational groups have been investigated but there is little agreement as to specific occupational exposures that may be implicated. Cohort studies done by Kaplan et al. (1986), Terreros et al. (1997) and a case control study reported by Mele et al. (1997) suggest benzene as a risk factor. Terreros et al. (1997), with only nine cases of myeloproliferative syndrome, reported an association with benzene exposure obtained from hospital case interview yielding on the basis of 9 cases an odds ratio (OR) of 46.6 (95% CI: 2.02,2761). Exposure to petroleum, of which benzene is a by-product, was associated with an increased risk of PV, ET (Kaplan, et.al., 1986). Quiroga (1981) reported a statistical association with exposure to petroleum. However, Pasqualetti et al. (1991) reported an elevated risk of haematological malignancies related to exposure to aromatic hydrocarbons, assessed as self-reported exposures to groups of chemicals (OR)=2.5 (95%1.39,3,32). This case control study assessed exposures via direct interview (59% of cases, 85% of controls) and the remaining questions obtained from relatives. From the list of occupational toxic substances gained from the interview, the researchers

then grouped these exposures into 21 at risk categories (Pasqualetti et al., 1991). No associations were found with myeloproliferative syndrome (MPS), (n=44, ORs not reported) (Pasqualetti et al., 1991). All of these studies were of limited size and did not use the molecular markers such as the JAK2 V617F mutation or the 46/1 haplotype that have significantly improved the identifying of true cases of MPNs (Seaman et al., 2009). Additonally, more advanced methods for occupational expsoure assesment are currently avaible, such as Job exposure matrix etc., that were not developed at the time of these studies. No association between MPNs and residential behaviors and presumed to be associated with exposure to aromatic and heterocyclic amines have been reported.

Genetic susceptibility to environmental exposures and PV, ET, PMF, and JAK2 V617F

Although functional genes that may modify the biological dose of a chemical mutagen have been studied for a wide range of cancers, no studies have investigated associations of these genotypes with MPNs. Susceptibility genotypes examined in these studies include, but are not limited to Cytochrome P450 superfamily (CYP genes), polymorphic N-acetyltransferase (NAT2), and glutathione S-transferase (GST) polymorphisms. These are all functional genes that encode enzymes on the pathway for biotransformation of genotoxic chemicals such as PAHs and benzene (Their, 2003). Because of the JAK2 V617F mutation and its role in MPNs, gene-environment interaction between genes that modify susceptibility to mutagenic chemicals and level of exposure to such chemicals is of particular interest.

Candidate gene and Single Nucleotide Polymorphisms (SNPs) selection overview

The families of genes that has been long studied and linked to increased cancer risk continue to be the CYP, GST, and NAT2 genes. One of the methods to estimate gene-

environment interactions is a candidate-gene approach using a biological pathway hypothesis as its framework (Hunter, 2005). The National Institute of Environmental Health Sciences Environmental Genome Project (NIEHS EGP) has already categorized environmentally sensitive genes (ESG) in the human genome. A total of 647 genes are included in the NIEHS EGP list in the following categories: cell structure, metabolism, DNA repair, cell cycle, cell division, cell signalling, gene expression, and homeostasis. This was the starting for our search for candidate genes, and specific polymorphisms to be used as a toxicological marker for susceptibility, including Cytochrome P450 superfamily (CYP genes), Polymorphic N-acetyltransferase (NAT2), and Glutathione S-transferase (GST) polymorphisms. The benefits of using Mendelian randomization are apparent when the genetic variants are functional SNPs and the biological function of the variant is known.

Garcia-Closas et al. (2011) reported on the use and utility of a novel tag SNP NAT2 (rs1495741) for the slow acetylator phenotype. In the past, a 7-SNP inferred phenotype had been used to assign individuals to the slow category, which was not always possible. Garcia-Closas reported a strong agreement between the rs1495741 tag SNP and established 7-SNP inferred NAT2 slow phenotype. They reported a sensitivity and specificity of 99 and 95% respectively for the tag SNP for slow (compared to intermediate/rapid) from 154 individuals with European background (Garcia-Closas, et.al, 2011).

Methods

Study design

The project used an unmatched case-control design. The study area was comprised of a tri-county area in Northeast Pennsylvania (Carbon, Luzern, and Schuylkill Counties). Cases included individuals with diagnoses of all MPNs combined, JAK2 V617F cases, and PV cases only. Controls were selected using random digit dialing sample stratified by age and county to obtain distributions that reflected that of the source counties. Cases and controls were evaluated for presence of the JAK2 V617F mutation (JAK2 V617F) and select genotypes susceptible to increases to genotoxic environmental toxins.

Participants

In 2007, an estimated 500,000 people resided in the tri-county region. People were eligible if they were born between 1921 and 1968 (42 and 89 years old). The restricted age group in the three counties was estimated to include 244,870 people. The study population was restricted by age to eliminate younger individuals not at risk for the MPNs. Since this study focuses on environmental exposures in the tri-county area, all subjects were required to reside in the tri-county area between 2000 and 2008.

Cases must have met the clinical criteria for a MPN and received a diagnosis between the years 2001 and 2010 to be eligible for inclusion in this study. The following case categorizations were used in this study for analysis: all MPNs, and confirmed MPN cases; JAK2 V617F mutation and confirmed JAK2 V617F mutation; PV only cases and confirmed PV only cases. Table 2.2 describes the case categorizations used in this study.

Cases were identified from the Pennsylvania Cancer Registry (PCR) using the International Classification of Diseases codes (ICD-O) for MPNs (codes M-9950/3, M-9962/3, M-9961/3; M-9931/3) as well as a cluster investigation published in 2009 for PV conducted by ATSDR. The second source was from a PV cluster investigation that included cases diagnosed between 2001 and 2006 in the three county area. We included cases diagnosed up to December 31, 2010.

All cases not previously confirmed by the ATSDR cluster investigation (Seaman et al., 2009) were evaluated by one of the two expert panels. Either hematologist's panel created by the Pennsylvania Department of Health or panel used in case ascertainment for a related study done by the University of Pittsburgh. Both panels used similar summary sheets, and a two thirds majority decision rule for case determination. The medial records of cases were summarized and both patients' and physicians' names were removed to facilitate impartial evaluation of the medical records. Full-case interviews only included surviving patients that consented to participate in the study. We also tried to contact eligible cases that had relocated from the tri-county area.

Controls were selected using random digit dialing. Individuals born from 1921-1968 and residing in these three counties were eligible. We selected a stratified random sample of controls on age and gender from the study area. The sample collected was similar to the population distribution documented by the from U.S. Census American Community Survey, 2008.

To participate in this genetic study subjects enrolled in the main study also had to consent to a blood draw. Genotype data were available for 57% of the cases and 62% of controls. The candidate gene-environment analysis was therefore based on information from approximately 31 MPNs (all MPNs combined) cases and 292 controls agreeing to the blood draw.

Data collection

Consent to compete the phone survey phase of the study was required to be eligible with an additional incentive $25 gift card to consent for a blood sample. A description of what consent entails was included at the bottom of the consent form, along with a yes/no box. As well, we offered a second incentive (a $25 gift card) for this phase, to encourage blood draw recruitment.

Blood draw was a one-time peripheral venipuncture of 25-30ml of blood. The maximum volume taken for samples was: (1)10ml for JAK2 V617F mutation testing, (2)10ml for gene susceptibility testing, (3)10ml for storage and possible future testing for biomarkers linked to MPN by other studies ongoing in the tri-county area (PV Partners studies). The blood samples were sent to at Columbia University Laboratory to perform genotyping. Mt. Sinai and Geisinger conducted the JAK2 V617F testing.

Genotyping procedure

Columbia University used the following steps to conduct genotyping: DNA was extracted from white blood cells by a standard salting-out protocol. DNA quality was assessed by absorption at 260 and 280 nm. Samples were aliquoted into 96 well plates for analysis. Genotyping for all selected SNPs, except rs#1048943 and rs#4646903, was carried out using the Illumina Bead Express platform that employs VeraCode technology (Illumina

San Diego CA). Rs#1048943 and rs#4646903 were genotyped by TaqMan™ assays (LifeTechnologies/Applied Biosystems, Carlsbad CA) in a 384 well plate format using an Applied Biosystems 7900 PCR system. About 7% of samples were run in duplicate for both SNP genotyping assays. Deletions in GSTM1 and GSTT1 were determined using TaqMan Copy Number Assays™ and RNase P as the control gene. Samples were run in triplicate and CopyCaller™ Software was used for determination of copy number.

Selection of environmentally sensitive genes

Because of the JAK2 V617F mutation and its role in PV-related outcomes, our interest in the mechanism of gene-environment interaction relates to genes that modify susceptibility to mutagenic chemicals. We considered only NIEHS genes and that were non-synonymous single nucleotide polymorphisms (nsSNPs), with a >5% minor allele frequency (MAF). From the full list of the 648 environmentally sensitive genes, 114 NIEHS metabolism genes were given initial consideration, predominantly, the Phase I and Phase II detoxifying enzyme classes. The functional gene groups of interest were genes that regulate enzymes of two categories: metabolism or detoxification. Although other functional categories may be associated with the chemicals of interest (e.g. PAHs), they are out of the scope of this research.

To facilitate the nsSNP selection, we needed an online database with genotyping information for each gene. The Genome Variation Server (GVS) includes detailed SNP information for nearly all EGP genes and is sponsored by Seattle SNPs Program for Genomic Applications (PGA), through the National Heart Lung and Blood Institute (NHLBI) (GVS,2011). Only 82 of the 114 had coding SNP data in the GVS database. We reviewed the 82 genes with nsSNP data for relevance to chemicals, and a mutagenic pathway of interest in the literature. We ultimately included 14 genes for this project, in

addition to the three genes without coding data, as shown in Table 3.1. Table 3.1 lists the mutagenic chemicals of interest. Based on this list, all but three chemicals (those being metals), are captured by the list of susceptibility genes.

In general the variants included in this analysis modify enzyme function, usually weakening the resultant gene product and thereby magnifying the effect of the xenobiotic substrate and are associated with a mutagenic chemical of interest (Dong et al., 2008). The genes we selected that may be associated with a mutagenic chemical (based on this biological pathway hypothesis) are AHR, CYP1A1, CYP1A2, CYP1B1, CYP2B6, GSTP1, GSTM1, GSTM3, GSTT1, NAT2, and NQO1 (see Table 3.1) (Genome Variation Server (GVS), 2010,2011) (Hung et al., 2003). One, SNP consistently did not perform well and was recommended by the Columbia University Laboratory to be omitted from our analysis.

The benefits of using Mendelian randomization are apparent when the genetic variants are functional SNPs and the biological function of the variant is known. Exploiting the assumptions of Mendelian randomizations that genotype and environmental exposures are independent and that risk of disease cannot vary with genotype without some environmental exposure. Gustasfon and Burstyn (2011) present this method as a practical approach to investigate gene-environment interaction.

Data analysis

Descriptive analysis was conducted on the characteristics of the study population. Logistic regression using SAS version 9.2 was used to estimate adjusted odds ratios (aOR) and associated 95% confidence intervals (CI). For genotypes with more than one functional SNP, we created and used dummy variables. A gene-only analysis (Burstyn et

al., 2009) used logistic regression to estimate unadjusted and adjusted odds ratios and 95% confidence intervals. The only covariates used were forced into the model. We controlled only for designed variables (sex, age, county) and conducted analysis of genotypes in the control population to demonstrate Hardy-Weinberg equilibrium in the genetic variants. We used the highest frequency of the homozygous genotype as the reference unless the literature indicated a different referent group. The reference groups used for each genotype are shown in Table 3.3. In addition, analysis of the number of deleterious SNPs was done, using logistic regression to estimate crude and adjusted ORs (aOR) with 95% confidence intervals.

Results

Demographics

Most (90%) of the cases were PV, 2 (7%) were diagnosed with ET, and 1 (>3%) was diagnosed with PMF. A greater proportion of cases were > age 65 (71 yrs. vs. 63 yrs.) and more likely to be male (55% vs. 45%) compared to controls but otherwise demographically similar (Table 3.5). The study population was overwhelmingly Caucasian (100% of cases and 99% of controls). None of the cases and few (2%) controls were of Jewish ancestry and all participants were born in the US. Two-thirds of the sample was currently married. More cases than controls were retired (65% versus 47%). These findings were similar across all case categories (Tables 3.4 and 3.5).

Susceptible genotypes associated with MPNs

Table 3.7 shows the main effects of the genetic polymorphisms. The crude estimates were very similar to the effect estimates that were adjusted for the design variables. The prevalence of CYP1A2, GSTA1, GSTM3, and NAT2 risk genotype in controls were 7%,

18%, 9% and 57%, respectfully, which was in agreement with the reported frequency in the literature (see Tables 3.1 and 3.6).

All consenting cases as a group were more likely to have the 46/1 haplotype (rs12340895), with an aOR of 5.3 (95% CI: 1.8, 15.7). For analysis restricted to JAK2 V617F and confirmed PV JAK2 V617F cases, respectfully, the effect estimates increased to aOR 10.2 (95% CI: 2.8, 37.9) and 10.8 (95% CI: 3.0, 38.2). All cases were also more likely to have the second 46/1 haplotype (rs12343867) (aOR of 4.3, 95% CI: 1.3, 14.7). Again, the effect estimates increased when the analysis was restricted to confirmed JAK2 V617F cases (aOR 7.9, 95% CI: 1.8, 34.9) or confirmed PV JAK2 V617F cases (aOR=9.6, 95% CI: 2.3, 39.8).

Having the CYP1A2 genotype increased the odds of MPNs by about four fold (aOR= 3.5, 95% CI: 1.1, 110). When case genotypes in the analysis were restricted to confirmed JAK2 V617F cases or PV JAK2 V617F cases, the effects estimates were similar (aOR 3.9, 95% CI: 1.1, 13.8 and aOR 3.7, 95% CI: 1.0, 13.3, respectively).

The GSTA1 genotype was associated with a 3 fold increase in effect estimates for all MPNs, with an aOR of 2.7 (95% CI: 1.0, 7.0) for confirmed JAK2 V617F cases only. The GSTM3 genotype increased the risk of MPNs: the aOR was 4.5 (95% CI: 1.2, 16.0) for all confirmed JAK2 V617F cases and was similar across all case categorizations.

The GSTM1 null genotype followed a similar trend of increasing ORs as the case categorizations were restricted to confirmed cases only, and JAK2 V617F only, although not all of these associations were statistically significant. We found a doubling of risk, aOR of 2.0 (95% CI: 0.9, 4.6), for risk of all MPNs, and an aOR of 2.2 (95% CI: 0.8, 6.0)

for all confirmed JAK2 V617F cases. When restricted to only confirmed PV JAK2 V617F cases, the aOR was 2.5 (95% CI: 0.9, 6.5).

We found that the adjusted OR of the NAT2 novel tag SNP (rs1495741) slow acetylator genotype compared to the wild type was three times greater among cases for all MPNs, (aOR=3.0, 95% CI:1.2,7.6). These associations were about the same magnitude yet understandably less precise when the analysis was limited to PV JAK2 V617F confirmed and JAK2 V617F case categorizations, (aOR=3.5, 95% CI:1.1,10.8) and (aOR=4.6, 95% CI:1.3,16.1) respectively. We were not able to test for all seven of the other NAT2 SNPs shown in the literature to determine slow acetylator phenotype, but we were able to test for four of them (shown in Table 3.7) that are consistent with slow phenotype. All but one (96.7%) case harbored at least two of the SNPs; AHR, CYP1E2, GSTM1, Tp53, GSTT1, GSTM3, CYP1A2, NAT2, and GSTA1 30/31 compared to 63.3 % in the control population (Table 3.8).

Discussion

After studying the main effects of 14 environmentally sensitive genes, we found that the *NAT2*, CYP1A2, *GSTA1*, and *GSTM3* variants were significantly associated with MPNs in this study sample ORs averaging between 3- to 5-fold. The prevalence of GSTA1, GSTM3, and NAT2 slow acetylator genotype in our controls was consistent with the reported frequency in the literature. While these results do not confirm a gene-environment interaction effect for any one specific chemical the findings encourage further explanation of the interaction hypothesis with respect to NAT2, GST, and CYP gene biological pathway and chemical exposures. These same genes appear to be implicated in presence of JAK2 V617F mutation also known to be implicated in aetiology of MPNs, especially PV.

To detect the potential for existence of gene-environment interaction, the main effects of genes were used in this analysis. By design we did not have a measure of exposures of interest and therefore could not estimate interactions or stratified effects directly. If we assume that genes alone do not cause MPNs, but can only act by modifying the toxicity of an environmental exposure, then testing the main effect of genes is an efficient way to generate evidence supporting qualitative gene-environment interaction (Burstyn et al., 2009, Gustafson & Burstyn, 2011). Based on this reasoning, we are essentially assuming there was a qualitative interaction and that there was no gene effect without exposure. A main gene effect without exposure is unlikely, based on the biological knowledge about the pathway for this disease. Therefore we have no reason to believe that MPNs (or any other disease for that matter) are caused exclusively by the genotypes under investigation.

In our study sample, we did not find consistent marginal effects from smoking and diet (data not shown) with risk of developing an MPN (Gross-Davis, Chapter 3, 2013). However, perhaps paradoxically, our findings suggest that specific genotypes that modify the toxicity of these exposures may play a role in MPNs. Because no relationship was seen in our previous work with cigarette smoking and MPNs, even though we had a detailed self-reported smoking history for all cases and controls and consequently are not concerned about exposure measurement error, despite the diminishing support for compounds present in tobacco smoke playing a role in MPNs. Lack of association of smoking with MPNs is consistent with all of the prior literature (Anderson et al., 2012). However, the literature on occupational exposures and chemical exposures are extremely vague with regard to details of what lifestyle exposure data was actually collected in these occupational cohorts as well as case control Kaplan, et.al.,(1986), Terreros et al. (1997), Zoloth et al., (1986), Johnson et al., (2010) and Pasqualetti et al. (1991).

Moore et al. (2011) reported an association of bladder cancer with NAT2 slow acetylator genotype and smoking intensity. They found that the NAT2 slow acetylator genotype and bladder cancer interaction differed with intensity of tobacco smoking but not ever, former or current smokers alone. Only slow acetylator types who smoked at least 40 cigarettes a day among ever smoked (aORs=1.82; CI: 95% 1.14, 2.91) and current smokers (aORs=3.16; CI: 95% 1.22, 8.19) compared to rapid acetylator showed significant risk. Since aromatic amines are detoxified by NAT2, interaction is biologically plausible. Thus, if carcinogens in tobacco smoke where implicated in MPNs, as they are in bladder cancer, then we should have detected the main effect of NAT2 in our study. However, absence of effect of smoking *per se* suggests that compounds not important to toxicity of tobacco smoke but affected by NAT2 should be scrutinized. The body of literature on chemical constituents of tobacco smoke is reported elsewhere and was not the focus on this study (Hecht,2003). Our analysis of smoking consistently refuted existence of positive association with very small numbers for cases for who were heavy smokers.

Other functional SNPs associated with benzene exposure were also explored in this analysis. These genotypes, CYP2E1, GSTM1, AHR, and GSTT1 also modify the biological dose of benzene and showed some interesting results worth reporting that support our MPN hypothesis. We observed a 2-fold increase for CYP2E1 across all case categorizations. A similar doubling of risk was found for individuals with the GSTM1 null genotype as well. These genotypes have been consistently reported in the literature in the pathway for the transformation of benzene (Lan, 2009). Looking at gene only effect the AHR and GSTT1 genotypes centered on the null with an aOR of 1.0-1.3 (95% CI: 0.4-3.8) for AHR and aOR=0.6-0.9 (95%CI: 0.2-2.6). In addition, the tumor

suppressor SNP, TP53 showed upwards of an 8-fold effect with an aOR=4.7-8.0 (95%CI: 1.2, 33) and has been reported to be involved with benzene hematopoietic stem-cell toxicity in mice (Hirabayashi, 2005).

Schnatter et al, updated three nested case–control studies in 2012 (Schnatter et al., 1996, Rushton et al., 1997, Glass et al., 2003). The resulting data was pooled to look at the risk of five specific lymphatic–hematopoietic subtypes and benzene exposure. The LH subtypes included were three types of leukemia -acute myeloid leukemia (AML), chronic myeloid leukemia (CML), and chronic lymphocytic leukemia (CLL) and – two types of myeloid neoplasm - myelodysplastic syndrome (MDS) and myeloproliferative disorders (MPD) (now referred to as MPNs). These nested studies were reviewed in 2010 by Miller, et al., and they concluded that these studies were similar in study design, and methods used to estimated benzene exposure. Any differences found were thought to be a result of how (cases) were studied and what data could be gathered for them, rather than any exposure assessments methods (Miller et al., 2010). This pooled study population consisted of 227 out of the original 370 cases and 1587 controls (Schnatter, et al., 2012). For the 29 cases of MDS, this pooled analysis showed a monotonic-dose response relationship with cumulative exposure to benzene (OR= 4.33 95% CI: 1.31, 14.3) but not for other subtypes including 30 MPD cases (OR=1.79, 95% CI: 0.68, 4.74) (Schnatter, et al., 2012).

Limitations

Our study suffered from a small number of cases. These MPNs are rare hematological malignancies with only a limited number of cases available for recruitment under ideal conditions. There were thirty one cases available for this study. Using the PCR as a source of cases, we did attempt to pull cases into our study that arose from this sample

population. However not optimal, our case categorizations allow us to look across all MPN outcomes and not be limited to only the number of incident cases of PV, ET, or PMF. By expanding the case definition to include individuals with the JAK2 V617F mutation, we increased our number of cases as much as possible, a for total of 31 MPN cases, which is reasonable with respect to other studies of MPN etiology, PV n=10 (Quiroga et al, 1981), ET= 133 (Mele et al., 1997, Falcetta et al., 2003) and myeloproliferative disorders n=53 (Pasqualetti, 1991, Terreros, 1997) and MPDs n=30 (Schnatter et al., 2012).

Although MPNs are now reported to the National Cancer Registry, MPNs only became a reportable disease in 2001 and in Pennsylvania only hospitals were required to report, so there is possible under reporting of these MPNs to the cancer registry. There were also changes in the diagnostic criteria by the World Health Organization (WHO) in 2001 and again in 2008 (Vakil&Tefferi, 2011). The 2008 diagnostic criteria included molecular as well as histological information for diagnosis, including the JAK2 V617F mutation (Vakil & Tefferi, 2011). Our ability to test for this JAK2 V617F mutation directly helped minimize case misclassification.

The University of Pittsburgh reported finding from a 2011 case ascertainment study in the tri county area both false reporting and underreporting in the tri-county area for PV and ET. They reported that 44% were true cases, 23% were false cases, 19% could not be determined due to a lack of information, and that 87% of the true cases were from the original PCR dataset. The reporting of the incidence and prevalence of these MPNs are not necessarily improving, and further monitoring of these diseases was recommended in this case ascertainment study (Buchanich & Mertz 2013). Since these MPNs often do not require hospitalization, reporting should be required and enforced at the

hematologist's office directly as this is not currently the case. The CDC has an ongoing effort to improve reporting of MPNs to cancer registries all over the country.

Our randomized selection of controls for the tri-county area attempted to minimize any concern over selection bias. Our response rate from MPN cases was small. We had a 1 to 9 ratio of cases and controls with no significant demographic differences observed between subjects who consented to genotyping and those who did not.

The latency period between possible exposures related to the development of an MPN is currently unknown, which complicates identification of potential risk factors when the timeframe of effective exposure is not clearly defined. However given this and other measurement challenges for exposures, assuming universal exposures and using a gene only analysis to detect gene-environmental interaction can be more reliable than assigning exposure groups to individuals.

Our study is vulnerable to false positive discovery do to "multiple comparisons" (Hunter et al., 2005). However, unlike GWAS studies, we started with 648 genes of interest and only examined 14 genes that met our *a priori* "plausible candidate". Our selective genotypes were functional SNPs, and did directly affect the enzymatic activity of the gene, influences biological pathways that affect metabolic activation/detoxification processes of metabolism for mutagenic chemicals. We did not correct for elevated type II error associations in applying a correction factor in the analysis.

Strengths

This study capitalized on the initial blood sample needed for JAK2 V617F mutation molecular testing to include additional genotyping for 14 genes using our biological pathway candidate gene approach. This study was not at risk from misclassification of

genotypes in our subjects as is often the case when testing for gene-environmental interaction directly using cross product of genetic marker and estimate of exposure in a statistical model or in stratified analysis (Hunter et al., 2005). As the misclassification error for DNA extraction are dependent on the genotype call rates, misclassification of genotypes was not a concern in this study because our call rate was high (>95%), and we only had one SNP that consistently did not perform well (Smith et al., 2011, Deitz et al., 2004).

Mendelian randomization is a strength here because we went to great lengths to identify functional SNPs that produced an effect that would modify the exposure of interest. While the exposure data is important, testing for known genotypes that are susceptible to environmental exposures allowed us to use the genotype as an instrument or proxy for the exposure. Another benefit of using the concept of Mendelian randomization minimizes the potential bias in exposure measurement (McKeigue et al., 2010; Lawlor, 2008; Verduijn et al., 2010).

Our analysis was done working under two main assumptions. First, that genotype and environmental exposures in this study population are independent. And second, that disease risk will not vary with genotype for subjects without environmental exposure. From this, we exploit the most generic form of Mendelian randomization that individuals receive a random allocation of alleles from their parents. Gustasfon and Burstyn (2011) present this method and concluded that when both of these assumptions are met, using data on genotype and disease jointly, with only knowledge of the prevalence of exposure without individual level data on exposure can be a practical approach to investigate gene-environment interaction. Through this gene-environment interaction approach, a

signal of increased risk will only result, if there is truly gene environment interaction (Smith et al., 2011).

None of the studied proxies of PAH exposure revealed associations, including smoking (Gross-Davis, Chapter 3, 2013). In addition, we did not see an association with the CYP1A1 SNP that would reveal a gene-environment interaction with PAHs. Our data does not support PAHs as a suspected exposure associated with development of an MPN.

In this application we use known biological pathways and candidate gene polymorphisms such as the GSTM3, CYP1A2 and NAT2 gene that modify environmental exposures due to increased ability to metabolize or decreased ability to detoxify the chemical. Since this if the first study to explore genetic polymorphisms and MPNs, this can help target future studies where only ecological environmental exposure data is available, with disease and genotype data.

Exposures other than smoking that are important to the pathways influenced by NAT2, GST, CYPs are of interest and for future investigations into the etiology of MPNs.

List of References

Anderson, L. A., Duncombe, A. S., Hughes, M., Mills, M. E., Wilson, J. C., & McMullin, M. F. (2012). Environmental, lifestyle, and familial/ethnic factors associated with myeloproliferative neoplasms. *Am J Hematol, 87*(2), 175-182.

Andrikovics, H., Nahajevszky, S., Koszarska, M., Meggyesi, N., Bors, A., Halm, G., . . . Tordai, A. (2010). JAK2 46/1 haplotype analysis in myeloproliferative neoplasms and acute myeloid leukemia. *Leukemia, 24*(10), 1809-1813.

Baris D, Karagas MR, Verrill C, Johnson A, Andrew AS, Marsit CJ, Schwenn M, Colt JS, Cherala S, Samanic C, Waddell R, Cantor KP, Schned A, Rothman N, Lubin J, Fraumeni JF Jr, Hoover RN, Kelsey KT, Silverman DT.
J Natl Cancer Inst. 2009 Nov 18;

Baxter EJ,Scott LM,Campbell PJ, et al. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet* 2005; **365**: 1054–1061.

Brockton N, Little J, Sharp L, Cotton SC. N-acetyltransferase poly- □morphisms and colorectal cancer: a HuGE review. Am J Epidemiol □2000;151:846 – 61.

Buchanich, J., Mertz, K., 2013. Updated and Expanded Study of Polycythemia Vera and Other Myeloproliferative Neoplasms in the Tri-County Area. FINAL REPORT to Pennsylvania Department of Health

Burstyn, I., Kim, H.-M., Yasui, Y., & Cherry, N. M. (2009). The virtues of a deliberately mis-specified disease model in demonstrating a gene-environment interaction. Occupational and Environmental Medicine, 66(6), 374-380.

Campbell, P. J., & Green, A. R. (2005). Management of polycythemia vera and essential thrombocythemia. *Hematology Am Soc Hematol Educ Program*, 201-208

Clapp, R. W., Jacobs, M. M., & Loechler, E. L. (2008). Environmental and occupational causes of cancer: new evidence 2005-2007. *Rev Environ Health, 23*(1), 1-37.

Deitz, A. C. et al. Impact of misclassification in genotype-exposure interaction studies: example of N-acetyltransferase 2 (NAT2), smoking, and bladder cancer. Cancer Epidemiol. Biomarkers Prev. 13, 1543–1546 (2004).

Dong, Linda M., Potter, John D., White, Emily, Ulrich, Cornelia M., Cardon, Lon R., & Peters, Ulrike. (2008). Genetic Susceptibility to Cancer. *JAMA: The Journal of the American Medical Association, 299*(20), 2423-2436.

Garcia-Closas, M., et al. (2013). "Common Genetic Polymorphisms Modify the Effect of Smoking on Absolute Risk of Bladder Cancer." Cancer Research 73(7): 2211-2220.

Genome Variation Server (GVS), . GVS was most recently updated February 26, 2010. The current version is 5.11). Retrieved accessed between September 2010 and February 2011, from < http://gvs.gs.washington.edu/GVS138/

Glass DC, Gray CN, Jolley DJ, et al. Leukemia risk associated with low-level benzene exposure. Epidemiology 2003. 14(5):569–577.

Green J, Banks E, Berrington A, Darby S, Deo H, Newton R. (2000) N-□acetyltransferase 2 and bladder cancer: an overview and consideration of the evidence for gene-environment interaction. Br J Cancer; 83:412–7.

Gross-Davis, CA. (2013)Environmental Etiology of Polycythemia Vera, Essential Thrombocythemia and Primary Myelofibrosis: A Case-Control Study in Northeast Pennsylvania disseration, Drexel Univeristy, School of Public Health

Gustafson, P.,Burstyn, I.(2011). Bayesian inference of gene–environment interaction from incomplete data: What happens when information on environment is disjoint from data on gene and disease? *Statistics in Medicine. Volume 30, Issue 8, pages 877–889,*

Hecht, Stephen S. "Tobacco carcinogens, their biomarkers and tobacco-induced cancer." Nature Reviews Cancer 3.10 (2003): 733-744.

Hein DW, Doll MA, Rustan TD, Gray K, Feng Y, Ferguson RJ, Grant DM. Metabolic activation and deactivation of arylamine carcinogens by recombinant human NAT1 and polymorphic NAT2 acetyltrans- ferases. Carcinogenesis 1993;14:1633–8

Hines LM, Stampfer MJ, Ma J et al. Genetic variation in alcohol dehydrogenase and the beneficial effect of moderate alcohol consumption on myocardial infarction. N Engl J Med 2001;344: 549–55.

Hirabayashi, Yoko. "p53-dependent gene profiling for reactive oxygen species after benzene inhalation: special reference to genes associated with cell cycle regulation." Chemico-biological interactions 153 (2005): 165-170.

Hunter, D. J. (2005). Gene–environment interactions in human diseases. *Nature Reviews Genetics*, *6*(4), 287-298.

Hung, R. J., Boffetta, P., Brockmoller, J., Butkiewicz, D., Cascorbi, I., Clapper, M. L., et al. (2003). CYP1A1 and GSTM1 genetic polymorphisms and lung cancer risk in Caucasian non-smokers: a pooled analysis. *Carcinogenesis, 24*(5), 875-882.

Jekarl, Dong Wook, et al. (2010) "JAK2 V617F mutation in myelodysplastic syndrome, myelodysplastic syndrome/myeloproliferative neoplasm, unclassifiable, refractory anemia with ring sideroblasts with thrombocytosis, and acute myeloid leukemia." The Korean journal of hematology 45.1: 46-50.

Johnson ES, Zhou Y, Lillian Yau C, et al. (2010)Mortality from malignant diseases-update of the Baltimore union poultry cohort. Cancer Causes Control;21:215–221.

Jones, Amy V., Chase, Andrew, Silver, Richard T., Oscier, David, Zoi, Katerina, Wang, Y. Lynn, . . . Cross, Nicholas C. P. (2009). JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nat Genet, 41*(4), 446-449.

Kaplan, S. D. (1986). Update of a mortality study of workers in petroleum refineries. *J Occup Med, 28*(7), 514-516.

Kilpivaara, Outi, Mukherjee, Semanti, Schram, Alison M., Wadleigh, Martha, Mullally, Ann, Ebert, Benjamin L., . . . Levine, Ross L. (2009). A germline JAK2 SNP is associated with predisposition to the development of JAK2V617F-positive myeloproliferative neoplasms. *Nat Genet, 41*(4), 455-459.

Kralovics R, Passamonti F, Buser AS, Teo S, Tiedt R, Passweg JR, Tichelli A, Cazzola M, Skoda RC. A gain-of-function mutation of JAK2 in myeloproliferative disorders. N. Engl. J. Med. 2005;352:1779–1790.

Kutti J,Ridell B. Epidemiology of the myeloproliferative disorders: Essential thrombocythaemia, polycythaemia vera and idiopathic myelofibrosis. *Pathol Biol* 2001; **49**: 164–166.

Lan Q, Zhang L, Shen M, Jo WJ, Vermeulen R, et al. 2009. Large-scale evaluation of candidate genes □identifies associations between DNA repair and genomic maintenance and development of benzene □hematotoxicity. Carcinogenesis 30:50–58

Landgren, O., Kristinsson, S. Y., Goldin, L. R., Caporaso, N. E., Blimark, C., Mellqvist, U. H., Turesson, I. (2009). Risk of plasma cell and lymphoproliferative disorders among

14621 first-degree relatives of 4458 patients with monoclonal gammopathy of undetermined significance in Sweden. *Blood, 114*(4), 791-795

Langabeer, S., Ni Ainle, F., Conneally, E., & Lawler, M. (2007). Incidence and significance of the <i>JAK2</i> V617F mutation in patients with chronic myeloproliferative disorders. *Irish Journal of Medical Science, 176*(2), 105-109.

Lawlor, D. A., Windmeijer, F., & Davey Smith, G. (2008). Is Mendelian randomization 'lost in translation?': Comments on 'Mendelian randomization equals instrumental variable analysis with genetic instruments' by Wehby et al. Statistics in Medicine, 27(15), 2750-2755.

McKeigue, P. M., Campbell, H., Wild, S., Vitart, V., Hayward, C., Rudan, I., et al. (2010). Bayesian methods for instrumental variable analysis with genetic instruments ('Mendelian randomization'): example with urate transporter SLC2A9 as an instrumental variable for effect of urate levels on metabolic syndrome. International Journal of Epidemiology, 39(3), 907-918.

Mele A, Visani G, Pulsoni A, et al. Risk factors for essential thrombocythe- mia—A case–control study. Cancer 1996;77:2157–2161.

Miller BG, Fransman W, Heederik D, Hurley JF, Kromhout H, Fitzsimons E. A review of the data quality and comparability of case-control studies of low-level exposure to benzene in the petroleum industry. Int Arch Occup Environ Health 2010 83(1):69–76.

Moliterno, A. R., Williams, D. M., Rogers, O., Isaacs, M. A., & Spivak, J. L. (2008). Phenotypic variability within the JAK2 V617F-positive MPD: roles of progenitor cell and neutrophil allele burdens. *Exp Hematol, 36*(11), 1480-1486.

Moore, L. E.,et al.,(2011).GSTM1 null and NAT2 slow acetylation genotypes, smoking intensity and bladder cancer risk: results from the New England bladder cancer study and NAT2 meta-analysis. Carcinogenesis. 32 (2), 182-1989.

Olcaydu, Damla, Harutyunyan, Ashot, Jager, Roland, Berg, Tiina, Gisslinger, Bettina, Pabinger, Ingrid, . . . Kralovics, Robert. (2009). A common JAK2 haplotype confers susceptibility to myeloproliferative neoplasms. *Nat Genet, 41*(4), 450-454.

Passamonti, Francesco, & Rumi, Elisa. (2009). Clinical relevance of JAK2 (V617F) mutant allele burden. *Haematologica, 94*(1), 7-10.

Pasqualetti P, Casale R, Colantonio D, Collacciani A. Occupational risk for hematological malignancies. Am J Hematol 1991;38:147–149.

Purdue, M. P., Hoppin, J. A., Blair, A., Dosemeci, M., & Alavanja, M. C. (2007). Occupational exposure to organochlorine insecticides and cancer incidence in the Agricultural Health Study. *Int J Cancer, 120*(3), 642-649.

Quiroga Micheo E, Calcagno EJ, Calabria SI, et al. Retrospective epidemio- logical study of hemopoietic system neoplasms in Argentina. Medicina 1981;41:187–200.

Rushton L, Romaniuk H. A case-control study to investigate the risk of leukaemia associated with exposure to benzene in petroleum marketing and distribution workers in the United Kingdom. Occup Environ Med 1997. 54(3):152–166.

Schnatter AR, Armstrong TW, Nicolich MJ, et al. (1996). Lymphohaematopoietic malignancies and quantitative estimates of exposure to benzene in Canadian petroleum distribution workers. Occup Environ Med. 53(11):773–781.

Schnatter, Robert Glass, Deborah C. Tang, Gong, Irons, Richard D. and Rushton,Lesley. (2012) Myelodysplastic (2012). Syndrome and Benzene Exposure Among Petroleum Workers: An International Pooled Analysis. JNCI J Natl Cancer Inst.

Seaman, V., Dearwent, S. M., Gable, D., Lewis, B., Metcalf, S., Orloff, K., . . . Cole, H. (2010). A multidisciplinary investigation of a polycythemia vera cancer cluster of unknown origin. *Int J Environ Res Public Health, 7*(3), 1139-1152.

Seaman, V., Jumaan, A., Yanni, E., Lewis, B., Neyer, J., Roda, P., . . . Hoffman, R. (2009). Use of molecular testing to identify a cluster of patients with polycythemia vera in eastern Pennsylvania. *Cancer Epidemiol Biomarkers Prev, 18*(2), 534-540.

Smith, G.D., Palmer,L., Burton,P., (2011). An Introduction to Genetic Epidemiology. pgs 140-141.by Policy Press, University of Bristol, Bristol, UK.

Spivak JL. (2010) Narrative review: Thrombocytosis, polycythemia vera, and JAK2 mutations:The phenotypic mimicry of chronic myelofibrosis. Ann Intern Med. 2010 Mar 2;152(5):300-6. Review.

Tefferi, A., Thiele, J., Orazi, A., Kvasnicka, H. M., Barbui, T., Hanson, C. A., . . . Vardiman, J. W. (2007). Proposals and rationale for revision of the World Health Organization diagnostic criteria for polycythemia vera, essential thrombocythemia, and primary myelofibrosis: recommendations from an ad hoc international expert panel. *Blood, 110*(4), 1092-1097.

Tefferi, A., & Vainchenker, W. (2011). Myeloproliferative neoplasms: molecular pathophysiology, essential clinical understanding, and treatment strategies. *J Clin Oncol, 29*(5), 573-582

Terreros MC, Apezteguia M, Slavutsky IR, Guimarey LM. Exposure to occupational and environmental risk factors in hematologic disorders. Neoplasia 1997;14:133–136.

Thier, Ricarda, et al. "Markers of genetic susceptibility in human environmental hygiene and toxicology: the role of selected CYP, NAT and GST genes." *International journal of hygiene and environmental health* 206.3 (2003): 149-171

Verduijn, M., Siegerink, B., Jager, K. J., Zoccali, C., & Dekker, F. W. (2010). Mendelian randomization: use of genetics to enable causal inference in observational studies. Nephrology Dialysis Transplantation, 25(5), 1394-1398.

Vijayakrishnan, J., & Houlston, R. S. (2010). Candidate gene association studies and risk of childhood acute lymphoblastic leukemia: a systematic review and meta-analysis. *Haematologica, 95*(8), 1405-1414.

Vakil, E., & Tefferi, A. (2011). BCR-ABL1--negative myeloproliferative neoplasms: a review of molecular biology, diagnosis, and treatment. *Clin Lymphoma Myeloma Leuk, 11 Suppl 1*, S37-45.

Wargovich, M. J., & Cunningham, J. E. (2003). Diet, individual responsiveness and cancer prevention. *J Nutr, 133*(7 Suppl), 2400S-2403S.

Zoloth SR, Michaels DM, Villalbi JR, Lacher M. (1986). Patterns of mortality among commercial pressmen. J Natl Cancer Inst. 1986 Jun;76(6):1047-51.

Table 3.1 Genes associated with a mutagenic chemical and Minor Allele Frequency in the study population

Genes	Functions (GVS)	Minor Allele Frequency (%)	Chemical Exposure
AHR	This gene encodes a ligand-activated transcription factor involved in the regulation of biological responses to planar aromatic hydrocarbons.	10 [1]	2,3,7,8-TCDD, benzo[a]pyrene 2,4'-DDT, Benzene
CYP1A1	CYP1A1 is also known as AHH (aryl hydrocarbon hydroxylase). It is involved in the metabolic activation of aromatic hydrocarbons (polycyclic aromatic hydrocarbons, PAH).	10 [1]	2,3,7,8-TCDD, sodium arsenite (NaAsO2), 2,4'-DDT, Aroclor-1260(weak), benzo[a]pyrene, benzo[k]fluoranthene , cadmium chloride
CYP1A2	This protein encoded by this gene localizes to the endoplasmic reticulum and its expression is induced by some polycyclic aromatic hydrocarbons (PAHs), some of which are found in cigarette smoke. The enzyme's endogenous substrate is unknown; however, it is able to metabolize some PAHs to carcinogenic intermediates. Other xenobiotic substrates for this enzyme include caffeine, aflatoxin B1, and acetaminophen.	24	2,3,7,8-TCDD, benzo[a]pyrene
CYP1B1	metabolizes procarcinogens such as polycyclic aromatic hydrocarbons and 17beta-estradiol.	44[1]	2,3,7,8-TCDD, benzo[a]pyrene
CYP2B6	This enzyme is known to metabolize some xenobiotics, such as the anti-cancer drugs cyclophosphamide and ifosphamide.	21 [1]	2,4'-DDT
CYP2E1	Inactivates a number of drugs and Xenobiotics and also bioactivates many xenobiotic substrates to their hepatotoxic or carcinogenic forms	6	Benzene, methylene chloride, Styrene, 1,1,1-trichloroethane, Trichloroethylene
GSTM1	This GST family member is a polymorphic gene encoding active, functionally different GSTP1 variant proteins that are thought to function in xenobiotic metabolism and play a role in susceptibility to cancer, and other diseases.	49 [1]	Styrene, Benzene
GSTM3	This GST family member is a polymorphic gene encoding active, functionally different GSTP1 variant proteins that are thought to function in xenobiotic metabolism and play a role in susceptibility to cancer, and other diseases.	32 [1]	
GSTT1	This GST family member is a polymorphic gene encoding active, functionally different GSTP1 variant proteins that are thought to function in xenobiotic metabolism and play a role in susceptibility to cancer, and other diseases.	20 [1]	benzo[a]pyrene ,methylene chloride, Styrene, Benzene

Genes	Functions (GVS)	Minor Allele Frequency (%)	Chemical Exposure
NAT2	Polymorphisms in this gene are responsible for the N-acetylation polymorphism in which human populations segregate into rapid, intermediate, and slow acetylator phenotypes. Polymorphisms in this gene are also associated with higher incidences of cancer and drug toxicity.	-	benzo[a]pyrene
NQO1	Mutations in this gene have been associated with tardive dyskinesia (TD), an increased risk of hematotoxicity after exposure to benzene, and susceptibility to various forms of cancer.	22[1]	2,3,7,8-TCDD, benzo[a]pyrene

Table 3.1 Genes associated with a mutagenic chemical and Minor Allele Frequency in the study population

Table 3.2 Case categorizations

Case definition[1]	Cases		JAK2 V617F mutation cases		controls	
	Total interviewed	Cases with genotype data	Total interviewed	cases with genotype data	Total interviewed	Controls with genotype data
All cases	55	31	25	22	473	292
Confirmed cases	41	27	25	22	473	
Confirmed PV cases	33	24	21	20	473	
Confirmed ET cases	7		3		473	
Confirmed PMF cases	1		1		473	

1. Cases confirmed by one of two expert panels, Pa Department of Health or University of Pittsburgh by reviewing medical records.

Table 3.3 Single Nucleotide polymorphisms (SNPs) of select genes with reference groups

Gene	SNP	Reference group(s)
AHR	rs2066853	GG
ARNT	rs12410394	GG or AG
CYP17A1	rs743572	AA
CYP19A1	rs700519	GG
CYP1A1	rs1048943	TT
	rs4646903	TT
CYP1A2	rs762551	AC or CC AC or AA AA
CYP1B1	rs1056836	GG
CYP2B6	rs3745274	AA
CYP2C9	rs1057910	AA
	rs1799853	GG
CYP2E1	rs2031920	GG
	rs2070673	TT
	rs6413432	AA
CYP3A4	rs2740574	AA
CYP3A5	rs776746	GG

Table 3.3 Single Nucleotide polymorphisms (SNPs) of select genes with reference groups (cont'd)

Gene	SNP	Reference group(s)
CYP4B1	rs2297810	GG
GSTA1	rs3957356	AG or GG AG or AA AA
GSTM1		>0
GSTM3	rs7483	GG
	rs1332018	AA
	rs1799735	GG
GSTP1	rs1695	AA
	rs1138272	GG
GSTT1		>0
GSTZ1	rs7972	GG
	rs1046428	GG
JAK2 V617F	rs12340895	GG
	rs12343867	AA

Table 3.3 Single Nucleotide polymorphisms (SNPs) of select genes with reference groups (cont'd)

Gene	SNP	Reference group(s)
NAT2	rs1041983	AG or GG
		AG or AA
		AA
	rs1495741	AG or GG
		AG or AA
		AA
	rs1799929	AG or GG
		AG or AA
		AA
	rs1799930	AG or GG
		AG or AA
		AA
	rs1801279	AG or GG
		AG or AA
		AA
	rs1801280	AG or GG
		AG or AA
		AA
NQO1	rs1131341	GG
	rs1800566	GG

Table 3.4 Study participants demographics in gene environment interaction analysis [1] (cont'd)

	All cases [2] and % frequencies n=31	Confirmed cases [2] and % frequencies n=27	Confirmed JAK2 V617F cases [2] and % frequencies n=22	Confirmed PV cases [2] and % frequencies n=24	Controls and % frequencies n=292
County					
Carbon	5 (16%)	4 (15%)	4 (18%)	4 (17%)	31 (11%)
Luzerne	15 (48%)	12 (44%)	10 (45%)	10 (42%)	177 (61%)
Schuylkill	11 (35%)	11 (41%)	8 (36%)	10 (42%)	84 (29%)
Age					
42-64	8 (26%)	7 (26%)	6 (27%)	5 (21%)	167 (57%)
65+	23 (74%)	20 (74%)	16 (73%)	19 (79%)	125 (43%)
Sex					
Male	17 (55%)	15 (56%)	12 (55%)	13 (54%)	118 (40%)
Female	14 (45%)	12 (44%)	10 (45%)	11 (46%)	174 (60%)
Race/ethnicity					
Non-Hisp. White	31 (100%)	27 (100%)	22 (100%)	24 (100%)	290 (99%)
Latino, Hispanic	0	0	0	0	2 (1%)
Non-Hisp. Black	0	0	0	0	0
Non-Hisp. Native American, Alaskan native	0	0	0	0	0
Non-Hisp. Asian or Pacific Islander	0	0	0	0	0
Non-Hisp. Other	0	0	0	0	0
Non-Hisp., multiple race	0	0	0	0	0
Country of origin					
USA	31 (100%)	27 (100%)	22 (100%)	24 (100%)	292 (100%)
Other	0	0	0	0	0

Table 3.4 Study participants demographics in gene environment interaction analysis [1] (cont'd)

	All cases [2] and % frequencies n=31	Confirmed cases [2] and % frequencies n=27	Confirmed JAK2 V617F cases [2] and % frequencies n=22	Confirmed PV cases [2] and % frequencies n=24	Controls and % frequencies n=292
State born in					
AK	0	0	0	0	1 (0%)
CA	1 (2%)	1 (2%)	0	0	1 (1%)
CT	1 (2%)	1 (2%)	1 (4%)	1 (3%)	0
FL	0	0	0	0	0
IL	0	0	0	0	1 (0%)
MA	1 (2%)	1 (2%)	1 (4%)	1 (3%)	0
MN	0	0	0	0	2 (0%)
MS	0	0	0	0	1 (0%)
NC	0	0	0	0	0
NH	0	0	0	0	1 (0%)
NJ	2 (4%)	2 (5%)	2 (8%)	2 (6%)	8 (3%)
NM	0	0	0	0	1 (0%)
NY	0	0	0	0	10 (3%)
OH	0	0	0	0	1 (0%)
OK	1 (2%)	0	0	0	0
PA	26 (87%)	22 (85%)	17 (80%)	19 (85%)	262 (89%)
SC	0	0	0	0	0
SD	0	0	0	0	1 (0%)
TN	0	0	0	0	0
TX	0	0	0	0	1 (0%)

Table 3.4 Study participants demographics in gene environment interaction analysis [1] (cont'd)

	All cases [2] and % frequencies n=31	Confirmed cases [2] and % frequencies n=27	Confirmed JAK2 V617F cases [2] and % frequencies n=22	Confirmed PV cases [2] and % frequencies n=24	Controls and % frequencies n=292
VA	0	0	0	0	1 (0%)
WV	1 (2%)	1 (2%)	1 (4%)	1 (3%)	0
non-US	0	0	0	0	0
Jewish ancestry					
Yes	0	0	0	0	6 (2%)
No	31 (100%)	27 (100%)	22 (100%)	24 (100%)	282 (98%)
Don't know	0	0	0	0	3
Missing	0	0	0	0	1
Marital status					
Married	22 (71%)	18 (67%)	15 (68%)	15 (63%)	183 (63%)
Widowed	7 (23%)	7 (26%)	5 (23%)	7 (29%)	47 (16%)
Currently single	2 (6%)	2 (7%)	2 (9%)	2 (8%)	62 (21%)
Education					
Less than high school	2 (6%)	2 (7%)	2 (9%)	2 (8%)	5 (2%)
High school / GED	16 (52%)	13 (48%)	11 (50%)	13 (54%)	106 (36%)
Some college	7 (23%)	7 (26%)	5 (23%)	5 (21%)	104 (36%)
Bachelors degree	3 (10%)	2 (7%)	2 (9%)	2 (8%)	36 (12%)
More than bachelors	3 (10%)	3 (11%)	2 (9%)	2 (8%)	41 (14%)

Table 3.4 Study participants demographics in gene environment interaction analysis [1] (cont'd)

	All cases [2] and % frequencies n=31	Confirmed cases [2] and % frequencies n=27	Confirmed JAK2 V617F cases [2] and % frequencies n=22	Confirmed PV cases [2] and % frequencies n=24	Controls and % frequencies n=292
Household income					
Less than $20,000	3 (11%)	3 (12%)	3 (15%)	3 (14%)	38 (14%)
$20,000 - $35,000	10 (36%)	9 (36%)	7 (35%)	9 (41%)	67 (24%)
$35,000 - $50,000	3 (11%)	2 (8%)	2 (10%)	2 (9%)	54 (20%)
$50,000 - $75,000	6 (21%)	6 (24%)	4 (20%)	6 (27%)	67 (24%)
More than $75,000	6 (21%)	5 (20%)	4 (20%)	2 (9%)	48 (18%)
Don't know	1	1	1	1	4
Refused	2	1	1	1	14
Current employment					
Employed for wages	9 (29%)	7 (26%)	6 (27%)	6 (25%)	101 (35%)
Self-employed	1 (3%)	1 (4%)	1 (5%)	1 (4%)	9 (3%)
Out of work for more than a year	1 (3%)	1 (4%)	1 (5%)	0	11 (4%)
Out of work for less than a year	0	0	0	0	6 (2%)
Homemaker	0	0	0	0	9 (3%)
Student	0	0	0	0	0
Retired	20 (65%)	18 (67%)	14 (64%)	17 (71%)	137 (47%)
Unable to work	0	0	0	0	19 (7%)

1. Self-reported socio economic factors, number and %
2. All cases include PV, ET and PMF, confirmed cases by expert panel include PV, ET, and PMF, confirmed JAK2 V617F mutation with PV, ET or PMF by expert panel, confirmed by expert panel PV only

Table 3.5 Study demographic in gene environment interaction logistic regression models [1]

	All cases OR [2] n=31	All cases aOR [3] n=31	Confirmed cases OR [2] n=27	Confirmed cases aOR [3] n=27	Confirmed JAK2 V617F OR [2] n=22	Confirmed JAK2 V617F aOR [3] n=22	Confirmed PV OR [2] n=24	Confirmed PV aOR [3] n=24
Marital status								
Married	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Widowed	1.5 (0.7, 3.0)	1.1 (0.5, 2.2)	1.6 (0.8, 3.3)	1.1 (0.5, 2.6)	1.7 (0.7, 4.3)	1.1 (0.4, 3.0)	1.7 (0.7, 3.8)	1.2 (0.5, 3.1)
Currently single	0.7 (0.3, 1.5)	0.7 (0.3, 1.6)	0.3 (0.1, 1.1)	0.3 (0.1, 1.1)	0.4 (0.1, 1.6)	0.4 (0.1, 1.7)	0.4 (0.1, 1.4)	0.4 (0.1, 1.6)
Education								
Less than high school	4.2 (1.1, 15.7)	2.8 (0.7, 10.9)	3.0 (0.6, 13.6)	1.9 (0.4, 9.3)	3.9 (0.6, 26.3)	2.4 (0.3, 17.4)	3.0 (0.5, 18.1)	1.8 (0.3, 12.1)
High school / GED	1.5 (0.6, 3.7)	1.2 (0.5, 3.0)	1.3 (0.5, 3.4)	1.0 (0.4, 2.7)	1.6 (0.4, 5.7)	1.2 (0.3, 4.3)	1.8 (0.6, 5.5)	1.4 (0.4, 4.4)
Some college	0.4 (0.1, 1.2)	0.4 (0.1, 1.1)	0.5 (0.2, 1.5)	0.4 (0.1, 1.3)	0.7 (0.2, 2.9)	0.6 (0.1, 2.5)	0.5 (0.1, 2.0)	0.4 (0.1, 1.7)
Bachelors degree	1.0 (0.3, 3.0)	0.9 (0.3, 2.9)	0.6 (0.1, 2.3)	0.6 (0.1, 2.4)	0.7 (0.1, 4.6)	0.7 (0.1, 4.7)	0.6 (0.1, 3.2)	0.6 (0.1, 3.3)
More than bachelors	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Household income								
Less than $20,000	1.0 (0.4, 2.7)	0.5 (0.2, 1.4)	0.8 (0.2, 2.5)	0.3 (0.1, 1.3)	0.6 (0.1, 2.6)	0.2 (0.1, 1.2)	2.1 (0.5, 9.0)	1.0 (0.2, 4.7)
$20,000 - $35,000	1.1 (0.5, 2.6)	0.6 (0.2, 1.5)	1.0 (0.4, 2.6)	0.5 (0.2, 1.5)	0.9 (0.3, 2.8)	0.4 (0.1, 1.4)	2.6 (0.7, 9.8)	1.3 (0.3, 5.5)
$35,000 - $50,000	0.7 (0.3, 1.9)	0.4 (0.2, 1.2)	0.6 (0.2, 1.8)	0.3 (0.1, 1.1)	0.3 (0.1, 1.5)	0.1 (0.0, 0.8)	1.2 (0.3, 5.6)	0.7 (0.1, 3.5)
$50,000 - $75,000	0.9 (0.3, 2.2)	0.7 (0.3, 1.8)	1.0 (0.4, 2.6)	0.8 (0.3, 2.2)	0.6 (0.2, 2.1)	0.4 (0.1, 1.7)	2.3 (0.6, 8.9)	1.7 (0.4, 7.0)
More than $75,000	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent

1. Self-reported socio economic factors
2 All cases include PV, ET and PMF, confirmed cases by expert panel include PV, ET, and PMF, confirmed JAK2 V617F mutation with PV, ET or PMF by expert panel, confirmed by expert panel PV only
3. Adjusted (for age, sex and county) Odds Ratios (95% confidence intervals)

Table 3.6. Select genotype frequencies in the study population for cases and controls

Susceptible genotype SNP and rs number[1]	All cases [2] % frequencies n=31	Confirmed cases [2] % frequencies n=27	Confirmed JAK2 V617F cases [2] % frequencies n=22	Confirmed PV cases [2] % frequencies n=24	Controls [3] % frequencies n=292
AHR rs2066853					
AG	6 (19%)	5 (19%)	5 (23%)	5 (21%)	59 (20%)
AA	0	0	0	0	1 (0 %)
GG	25 (81%)	22 (81%)	17 (77%)	19 (79%)	230 (79%)
Missing	0	0	0	0	2
ARNT rs12410394					
GG	8 (26%)	7 (26%)	7 (32%)	6 (25%)	122 (42%)
AA	2 (6%)	1 (4%)	0	1 (4%)	33 (11%)
AG	21 (68%)	19 (70%)	15 (68%)	17 (71%)	135 (47%)
Missing	0	0	0	0	2
ARNT rs12410394					
AA	2 (6%)	1 (4%)	0	1 (4%)	33 (11%)
AG	21 (68%)	19 (70%)	15 (68%)	17 (71%)	135 (47%)
GG	8 (26%)	7 (26%)	7 (32%)	6 (25%)	122 (42%)
Missing	0	0	0	0	2
CYP17A1 rs743572					
AG	13 (42%)	12 (44%)	11 (50%)	11 (46%)	140 (48%)
GG	5 (16%)	4 (15%)	3 (14%)	4 (17%)	44 (15%)
AA	13 (42%)	11 (41%)	8 (36%)	9 (38%)	106 (37%)
Missing	0	0	0	0	2

Table 3.6. Select genotype frequencies in the study population for cases and controls (cont'd)

Susceptible genotype SNP and rs number[1]	All cases [2] % frequencies n=31	Confirmed cases [2] % frequencies n=27	Confirmed JAK2 V617F cases [2] % frequencies n=22	Confirmed PV cases [2] % frequencies n=24	Controls [3] % frequencies n=292
CYP19A1 rs700519					
AG	2 (6%)	2 (7%)	1 (5%)	1 (4%)	21 (7%)
GG	29 (94%)	25 (93%)	21 (95%)	23 (96%)	267 (93%)
Missing	0	0	0	0	4
CYP1A2 rs762551					
AA	12 (39%)	11 (41%)	9 (41%)	10 (42%)	145 (50%)
AC	14 (45%)	11 (41%)	9 (41%)	10 (42%)	128 (44%)
CC	5 (16%)	5 (19%)	4 (18%)	4 (17%)	19 (7%)
CYP1A2 rs762551					
CC	5 (16%)	5 (19%)	4 (18%)	4 (17%)	19 (7%)
AC	14 (45%)	11 (41%)	9 (41%)	10 (42%)	128 (44%)
AA	12 (39%)	11 (41%)	9 (41%)	10 (42%)	145 (50%)
CYP1A2 rs762551					
AC or CC	19 (61%)	16 (59%)	13 (59%)	14 (58%)	147 (50%)
AA	12 (39%)	11 (41%)	9 (41%)	10 (42%)	145 (50%)
CYP1B1 rs1056836					
CC	10 (32%)	9 (33%)	7 (32%)	7 (29%)	52 (18%)
CG	12 (39%)	9 (33%)	6 (27%)	8 (33%)	148 (51%)
GG	9 (29%)	9 (33%)	9 (41%)	9 (38%)	91 (31%)
Missing	0	0	0	0	1

Table 3.6. Select genotype frequencies in the study population for cases and controls (cont'd)

Susceptible genotype SNP and rs number[1]	All cases [2] % frequencies n=31	Confirmed cases [2] % frequencies n=27	Confirmed JAK2 V617F cases [2] % frequencies n=22	Confirmed PV cases [2] % frequencies n=24	Controls [3] % frequencies n=292
CYP2B6					
rs3745274					
AC	21 (68%)	17 (63%)	14 (64%)	16 (67%)	164 (56%)
AA	10 (32%)	10 (37%)	8 (36%)	8 (33%)	128 (44%)
CYP2C9					
rs1057910					
AC	5 (16%)	4 (15%)	3 (14%)	4 (17%)	32 (11%)
AA	26 (84%)	23 (85%)	19 (86%)	20 (83%)	257 (89%)
Missing	0	0	0	0	3
rs1799853					
AG	6 (19%)	6 (22%)	5 (23%)	4 (17%)	64 (22%)
GG	25 (81%)	21 (78%)	17 (77%)	20 (83%)	225 (78%)
Missing	0	0	0	0	3
CYP2E1					
rs2031920					
AG	2 (6%)	2 (7%)	2 (9%)	2 (8%)	12 (4%)
GG	29 (94%)	25 (93%)	20 (91%)	22 (92%)	275 (96%)
Missing	0	0	0	0	5
rs2070673					
AT	6 (19%)	5 (19%)	4 (18%)	4 (17%)	74 (26%)
AA	0	0	0	0	9 (3%)
TT	25 (81%)	22 (81%)	18 (82%)	20 (83%)	206 (71%)
Missing	0	0	0	0	3

Table 3.6. Select genotype frequencies in the study population for cases and controls (cont'd)

Susceptible genotype SNP and rs number[1]		All cases [2] % frequencies n=31	Confirmed cases [2] % frequencies n=27	Confirmed JAK2 V617F cases [2] % frequencies n=22	Confirmed PV cases [2] % frequencies n=24	Controls [3] % frequencies n=292
rs6413432						
	AT	7 (23%)	6 (22%)	6 (27%)	6 (25%)	48 (17%)
	TT	0	0	0	0	2 (1%)
	AA	24 (77%)	21 (78%)	16 (73%)	18 (75%)	240 (83%)
	Missing	0	0	0	0	2
CYP3A4 rs2740574						
	AG	0	0	0	0	14 (5%)
	AA	30 (100%)	26 (100%)	22 (100%)	23 (100%)	277 (95%)
	Missing	1	1	0	1	1
CYP3A5 rs776746						
	AA	1 (3%)	1 (4%)	1 (5%)	1 (4%)	1 (0%)
	AG	3 (10%)	3 (11%)	2 (9%)	3 (13%)	38 (13%)
	GG	27 (87%)	23 (85%)	19 (86%)	20 (83%)	252 (87%)
	Missing	0	0	0	0	1
CYP4B1 rs2297810						
	AG	10 (32%)	9 (33%)	7 (32%)	9 (38%)	74 (26%)
	AA	0	0	0	0	3 (1%)
	GG	21 (68%)	18 (67%)	15 (68%)	15 (63%)	212 (73%)
	Missing	0	0	0	0	3

Table 3.6. Select genotype frequencies in the study population for cases and controls (cont'd)

Susceptible genotype SNP and rs number[1]	All cases [2] % frequencies n=31	Confirmed cases [2] % frequencies n=27	Confirmed JAK2 V617F cases [2] % frequencies n=22	Confirmed PV cases [2] % frequencies n=24	Controls [3] % frequencies n=292
GSTA1 rs3957356					
AA	10 (32%)	9 (33%)	9 (41%)	9 (38%)	53 (18%)
AG	14 (45%)	11 (41%)	7 (32%)	9 (38%)	148 (51%)
GG	7 (23%)	7 26%	6 (27%	6 (25%)	91 (31%)
GSTA1 rs3957356					
GG	7 (23%)	7 (26%)	6 (27%)	6 (25%)	91 (31%)
AG	14 (45%)	11 (41%)	7 (32%)	9 (38%)	148 (51%)
AA	10 (32%)	9 (33%)	9 (41%)	9 (38%)	53 (18%)
GSTA1 rs3957356					
AG or GG	21 (68%)	18 (67%)	13 (59%)	15 (63%)	239 (82%)
AA	10 (32%)	9 (33%)	9 (41%)	9 (38%)	53 (18%)
GSTM1					
Undetermined	0	0	0	0	1 (0%)
0	22 (71%)	20 (74%)	16 (73%)	18 (75%)	155 (53%)
1	9 (29%)	7 (26%)	6 (27%)	6 (25%)	116 (40%)
2	0	0	0	0	20 (7%)
GSTM3 rs7483					
AA	6 (19%)	6 (22%)	5 (23%)	5 (21%)	26 (9%)
AG	12 (39%)	11 (41%)	10 (45%)	9 (38%)	124 (43%)
GG	13 (42%)	10 (37%)	7 (32%)	10 (42%)	137 (48%)
Missing	0	0	0	0	5

Table 3.6. Select genotype frequencies in the study population for cases and controls (cont'd)

Susceptible genotype SNP and rs number[1]	All cases [2] % frequencies n=31	Confirmed cases [2] % frequencies n=27	Confirmed JAK2 V617F cases [2] % frequencies n=22	Confirmed PV cases [2] % frequencies n=24	Controls [3] % frequencies n=292
rs1332018					
CC	8 (26%)	7 (26%)	5 (23%)	7 (29%)	45 (16%)
AC	9 (29%)	7 (26%)	6 (27%)	7 (29%)	156 (54%)
AA	14 (45%)	13 (48%)	11 (50%)	10 (42%)	89 (31%)
Missing	0	0	0	0	2
rs1799735					
TG	5 (16%)	4 (15%)	3 (14%)	4 (17%)	81 (28%)
TT	0	0	0	0	8 (3%)
GG	26 (84%)	23 (85%)	19 (86%)	20 (83%)	200 (69%)
Missing	0	0	0	0	3
GSTM3					
rs7483					
AA	6 (19%)	6 (22%)	5 (23%)	5 (21%)	26 (9%)
AG	12 (39%)	11 (41%)	10 (45%)	9 (38%)	124 (43%)
GG	13 (42%)	10 (37%)	7 (32%)	10 (42%)	137 (48%)
Missing	0	0	0	0	5
GSTP1					
rs1695					
AG	15 (48%)	13 (48%)	10 (45%)	11 (46%)	135 (47%)
GG	0	0	0	0	33 (11%)
AA	16 (52%)	14 (52%)	12 (55%)	13 (54%)	122 (42%)
Missing	0	0	0	0	2

Table 3.6. Select genotype frequencies in the study population for cases and controls (cont'd)

Susceptible genotype SNP and rs number[1]	All cases [2] % frequencies n=31	Confirmed cases [2] % frequencies n=27	Confirmed JAK2 V617F cases [2] % frequencies n=22	Confirmed PV cases [2] % frequencies n=24	Controls [3] % frequencies n=292
rs1138272					
AA	4 (13%)	4 (15%)	3 (14%)	3 (13%	(35 12%)
AG	1 (3%)	1 (4%)	0	1 (4%)	30 (10%)
GG	26 (84%)	22 (81%)	19 (86%)	20 (83%)	227 78%
GSTT1					
3	0	0	0	0	1 (0%)
Undetermined	0	0	0	0	1 (0%)
0	5 (16%)	4 (15%)	2 (9%)	3 (13%)	46 (16%)
1	25 (81%)	22 (81%)	20 (91%)	21 (88%)	197 (67%)
2	1 (3%)	1 (4%)	0	0	47 (16%)
GSTZ1					
rs7972					
AA	1 (3%)	1 (4%)	1 (5%)	1 (4%)	6 (2%)
AG	5 (16%)	5 (19%)	5 (23%)	5 (21%)	44 (15%)
GG	25 (81%)	21 (78%)	16 (73%)	18 (75%)	242 (83%)
rs1046428					
AG	8 (27%)	6 (23%)	4 (19%)	5 (22%)	98 (34%)
AA	0	0	0	0	9 (3%)
GG	22 (73%)	20 (77%)	17 (81%)	18 (78%)	184 (63%)
Missing	1	1	1	1	1

Table 3.6. Select genotype frequencies in the study population for cases and controls (cont'd)

Susceptible genotype SNP and rs number[1]	All cases [2] % frequencies n=31	Confirmed cases [2] % frequencies n=27	Confirmed JAK2 V617F cases [2] % frequencies n=22	Confirmed PV cases [2] % frequencies n=24	Controls [3] % frequencies n=292
NAT2					
rs1041983					
AA	2 (6%)	2 (7%)	2 (9%)	1 (4%)	35 (12%)
AG	15 (48%)	12 (44%)	8 (36%)	11 (46%)	116 (40%)
GG	14 (45%)	13 (48%)	12 (55%)	12 (50%)	140 (48%)
Missing	0	0	0	0	1
rs1495741					
AA	25 (81%)	22 (81%)	19 (86%)	20 (83%)	164 (57%)
AG	5 (16%)	4 (15%)	2 (9%)	3 (13%)	104 (36%)
GG	1 (3%)	1 (4%)	1 (5%)	1 (4%)	22 (8%)
Missing	0	0	0	0	2
rs1799929					
AA	11 (35%)	10 (37%)	10 (45%)	9 (38%)	56 (19%)
AG	10 (32%)	8 (30%)	5 (23%)	8 (33%)	124 (42%)
GG	10 (32%)	9 (33%)	7 (32%)	7 (29%)	112 (38%)
rs1799930					
AA	1 (3%)	1 (4%)	1 (5%)	0	29 (10%)
AG	14 (47%)	12 (44%)	8 (36%)	11 (46%)	114 (39%)
GG	15 (50%)	14 (52%)	13 (59%)	13 (54%)	148 (51%)
Missing	1	0	0	0	1

Table 3.6. Select genotype frequencies in the study population for cases and controls (cont'd)

Susceptible genotype SNP and rs number[1]		All cases [2] % frequencies n=31	Confirmed cases [2] % frequencies n=27	Confirmed JAK2 V617F cases [2] % frequencies n=22	Confirmed PV cases [2] % frequencies n=24	Controls [3] % frequencies n=292
rs1801279						
	AA					
	AG					
	GG	31 (100%)	27 (100%)	22 (100%)	24 (100%)	291 (100%)
	Missing	0	0	0	0	1
rs1801280						
	AA	8 (26%)	7 (26%)	5 (23%)	5 (21%)	106 (36%)
	AG	11 (35%)	9 (33%)	6 (27%)	9 (38%)	127 (44%)
	GG	12 (39%)	11 (41%)	11 (50%)	10 (42%)	58 (20%)
	Missing	0	0	0	0	1
NAT2 rs1041983						
	GG	14 (45%)	13 (48%)	12 (55%)	12 (50%)	140 (48%)
	AG	15 (48%)	12 (44%)	8 (36%)	11 (46%)	116 (40%)
	AA	2 (6%)	2 (7%)	2 (9%)	1 (4%)	35 (12%)
	Missing	0	0	0	0	1
rs1495741						
	GG	1 (3%)	1 (4%)	1 (5%)	1 (4%)	22 (8%)
	AG	5 (16%)	4 (15%)	2 (9%)	3 (13%)	104 (36%)
	AA	25 (81%)	22 (81%)	19 (86%)	20 (83%)	164 (57%)
	Missing	0	0	0	0	2
rs1799929						
	GG	10 (32%)	9 (33%)	7 (32%)	7 (29%)	112 (38%)
	AG	10 (32%)	8 (30%)	5 (23%)	8 (33%)	124 (42%)
	AA	11 (35%)	10 (37%)	10 (45%)	9 (38%)	56 (19%)

Table 3.6. Select genotype frequencies in the study population for cases and controls (cont'd)

Susceptible genotype SNP and rs number[1]	All cases [2] % frequencies n=31	Confirmed cases [2] % frequencies n=27	Confirmed JAK2 V617F cases [2] % frequencies n=22	Confirmed PV cases [2] % frequencies n=24	Controls [3] % frequencies n=292
rs1799930					
GG	15 (50%)	14 (52%)	13 (59%)	13 (54%)	148 (51%)
AG	14 (47%)	12 (44%)	8 (36%)	11 (46%)	114 (39%)
AA	1 (3%)	1 (4%)	1 (5%)	0	29 (10%)
Missing	1	0	0	0	1
rs1801279					
GG	31 (100%)	27 (100%)	22 (100%)	24 (100%)	291 (100%)
AG					
AA					
Missing	0	0	0	0	1
rs1801280					
GG	12 (39%)	11 (41%)	11 (50%)	10 (42%)	58 (20%)
AG	11 (35%)	9 (33%)	6 (27%)	9 (38%)	127 (44%)
AA	8 (26%)	7 (26%)	5 (23%)	5 (21%)	106 (36%)
Missing	0	0	0	0	1
NAT2					
rs1041983					
AG or GG	29 (94%)	25 (93%)	20 (91%)	23 (96%)	256 (88%)
AA	2 (6%)	2 (7%)	2 (9%)	1 (4%)	35 (12%)
Missing	0	0	0	0	1
rs1495741					
AG or GG	6 (19%)	5 (19%)	3 (14%)	4 (17%)	126 (43%)
AA	25 (81%)	22 (81%)	19 (86%)	20 (83%)	164 (57%)
Missing	0	0	0	0	2

Table 3.6. Select genotype frequencies in the study population for cases and controls (cont'd)

Susceptible genotype SNP and rs number[1]	All cases [2] % frequencies n=31	Confirmed cases [2] % frequencies n=27	Confirmed JAK2 V617F cases [2] % frequencies n=22	Confirmed PV cases [2] % frequencies n=24	Controls [3] % frequencies n=292
rs1799929					
AG or GG	20 (65%)	17 (63%)	12 (55%)	15 (63%)	236 (81%)
AA	11 (35%)	10 (37%)	10 (45%)	9 (38%)	56 (19%)
rs1799930					
AG or GG	29 (97%)	26 (96%)	21 (95%)	24 (100%)	262 (90%)
AA	1 3%	1 (4%)	1 (5%)	0	29 (10%)
Missing	1	0	0	0	1
rs1801279					
AG or GG	31 (100%)	27 (100%)	22 (100%)	24 (100%)	291 (100%)
AA					
Missing	0	0	0	0	1
rs1801280					
AG or GG	23 (74%)	20 (74%)	17 (77%)	19 (79%)	185 (64%)
AA	8 (26%)	7 (26%)	5 (23%)	5 (21%)	106 (36%)
Missing	0	0	0	0	1
Gene = NQO1					
rs1131341					
AG	3 (10%)	3 (11%)	3 (14%)	2 (8%)	17 (6%)
GG	28 (90%)	24 (89%)	19 (86%)	22 (92%)	275 (94%)
rs1800566					
AA	1 (3%)	1 (4%)	0	1 (4%)	12 (4%)
AG	13 (42%)	12 (44%)	10 (45%)	10 (42%)	87 (30%)
GG	17 (55%)	14 (52%)	12 (55%)	13 (54%)	193 (66%)

Table 3.6. Select genotype frequencies in the study population for cases and controls (cont'd)

Susceptible genotype SNP and rs number[1]	All cases [2] % frequencies n=31	Confirmed cases [2] % frequencies n=27	Confirmed JAK2 V617F cases [2] % frequencies n=22	Confirmed PV cases [2] % frequencies n=24	Controls [3] % frequencies n=292
CYP1A1					
rs1048943					
CT	3 (10%)	3 (12%)	3 (14%)	2 (9%)	18 (6%)
TT	27 (90%)	23 (88%)	19 (86%)	21 (91%)	269 (94%)
Missing	1	1	0	1	5
rs4646903					
CC	1 (3%)	1 (4%)	0	1 (4%)	3 (1%)
CT	8 (27%)	6 (23%)	5 (24%)	5 (22%)	46 (16%)
TT	21 (70%)	19 (73%)	16 (76%)	17 (74%)	239 (83%)
Missing	1	1	1	1	4
Tp53					
rs1042522					
GG	-	4 (15%)	4 (18%)	4 (17%)	12 (4%)
CG	-	10 (37%)	9 (41%)	9 (38%)	117 (40%)
CC	-	13 (48%)	9 (41%)	11 (46%)	160 (55%)

1. Susceptible genes to mutagenic chemicals and Single Nucleotide Polymorphisms (SNPs) with rs number
2. All cases include PV, ET and PMF, confirmed cases by expert panel include PV, ET, and PMF, confirmed JAK2 V617F mutation with PV, ET or PMF by expert panel, confirmed by expert panel PV only
3. Controls who also consented to provide a blood sample for genotyping

Table 3.7 Select environmentally sensitive genotype logistic regression mod

	All cases OR [1] n=31	All cases aOR [2] n=31	Confirmed cases OR [1] n=27	Confirmed cases aOR [2] n=27	Confirmed JAK2 V617F OR [1] n=22	Confirmed JAK2 V617F aOR [2] n=22	Confirmed PV OR [1] n=24	Confirmed PV aOR [2] n=24
AHR rs2066853								
AG	0.9 (0.4, 2.4)	1.0 (0.4, 2.7)	0.9 (0.3, 2.4)	1.0 (0.4, 2.9)	1.1 (0.4, 3.2)	1.3 (0.4, 3.8)	1.0 (0.4, 2.9)	1.2 (0.4, 3.5)
AA	--	--	--	--	--	--	--	--
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
ARNT rs12410394								
GG	0.4 (0.2, 1.0)	0.4 (0.2, 1.0)	0.4 (0.2, 1.0)	0.4 (0.2, 1.0)	0.5 (0.2, 1.3)	0.5 (0.2, 1.3)	0.4 (0.1, 1.0)	0.4 (0.1, 1.0)
AA	0.4 (0.1, 1.7)	0.4 (0.1, 1.9)	0.2 (0.0, 1.7)	0.2 (0.0, 1.8)	--	--	0.2 (0.0, 1.9)	0.2 (0.0, 2.0)
AG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
ARNT rs12410394								
AA	0.5 (0.1, 2.4)	0.6 (0.1, 2.7)	0.3 (0.0, 2.3)	0.3 (0.0, 2.4)	--	--	0.3 (0.0, 2.6)	0.3 (0.0, 2.8)
AG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent

Missing								
ARNT rs12410394								
GG	0.5 (0.2, 1.1)	0.5 (0.2, 1.1)	0.5 (0.2, 1.2)	0.5 (0.2, 1.2)	0.6 (0.3, 1.6)	0.6 (0.2, 1.6)	0.5 (0.2, 1.2)	0.4 (0.2, 1.2)
AG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
ARNT rs12410394								
AG or GG	1.9 (0.4, 8.2)	1.7 (0.4, 7.9)	3.3 (0.4, 25.4)	3.2 (0.4, 25.6)	--	--	3.0 (0.4, 22.6)	2.9 (0.4, 23.3)
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
CYP17A1 rs743572								
AG	0.8 (0.3, 1.7)	0.8 (0.4, 1.9)	0.8 (0.4, 1.9)	0.9 (0.4, 2.2)	1.0 (0.4, 2.7)	1.1 (0.4, 2.9)	0.9 (0.4, 2.3)	1.0 (0.4, 2.7)
GG	0.9 (0.3, 2.8)	0.8 (0.2, 2.4)	0.9 (0.3, 2.9)	0.7 (0.2, 2.5)	0.9 (0.2, 3.6)	0.7 (0.2, 3.0)	1.1 (0.3, 3.7)	0.8 (0.2, 3.1)
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
CYP19A1 rs700519								
AG	0.9 (0.2, 3.9)	1.0 (0.2, 4.9)	1.0 (0.2, 4.6)	1.2 (0.3, 5.8)	0.6 (0.1, 4.7)	0.7 (0.1, 5.5)	0.6 (0.1, 4.3)	0.7 (0.1, 5.6)

GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
CYP1A2 rs762551								
AA	0.6 (0.3, 1.4)	0.7 (0.3, 1.5)	0.7 (0.3, 1.6)	0.8 (0.3, 1.7)	0.7 (0.3, 1.7)	0.8 (0.3, 1.9)	0.7 (0.3, 1.7)	0.8 (0.3, 1.9)
AC	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
CC	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
CYP1A2 rs762551								
CC	2.8 (1.0, 8.0)	3.5 (1.1, 11.0)	3.3 (1.1, 9.6)	4.1 (1.3, 13.2)	3.2 (1.0, 10.4)	3.9 (1.1, 13.8)	2.9 (0.9, 9.3)	3.7 (1.0, 13.3)
AC	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
CYP1A2 rs762551								
AC or CC	1.6 (0.7, 3.3)	1.4 (0.7, 3.2)	1.4 (0.6, 3.2)	1.3 (0.6, 3.0)	1.4 (0.6, 3.4)	1.3 (0.5, 3.2)	1.4 (0.6, 3.2)	1.2 (0.5, 3.0)
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
CYP1B1 rs1056836								
CC	1.9 (0.7, 5.1)	2.1 (0.8, 5.7)	1.8 (0.7, 4.7)	1.9 (0.7, 5.4)	1.4 (0.5, 3.9)	1.5 (0.5, 4.4)	1.4 (0.5, 3.9)	1.5 (0.5, 4.5)
CG	0.8 (0.3, 2.0)	0.9 (0.3, 2.2)	0.6 (0.2, 1.6)	0.7 (0.2, 1.8)	0.4 (0.1, 1.2)	0.4 (0.1, 1.3)	0.5 (0.2, 1.5)	0.6 (0.2, 1.6)
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								

CYP2B6								
rs3745274								
AC	1.6 (0.7, 3.6)	1.4 (0.6, 3.1)	1.3 (0.6, 3.0)	1.1 (0.5, 2.6)	1.4 (0.6, 3.4)	1.2 (0.5, 3.0)	1.6 (0.6, 3.8)	1.3 (0.5, 3.2)
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
CYP2C9								
rs1057910								
AC	1.5 (0.6, 4.3)	1.4 (0.5, 4.2)	1.4 (0.5, 4.3)	1.4 (0.4, 4.5)	1.3 (0.4, 4.5)	1.2 (0.3, 4.5)	1.6 (0.5, 5.0)	1.6 (0.5, 5.3)
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs1799853								
AG	0.8 (0.3, 2.1)	0.7 (0.3, 2.0)	1.0 (0.4, 2.6)	0.9 (0.3, 2.5)	1.0 (0.4, 2.9)	0.9 (0.3, 2.6)	0.7 (0.2, 2.1)	0.6 (0.2, 1.9)
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
CYP2E1								
rs2031920								
AG	1.6 (0.3, 7.4)	1.7 (0.3, 8.2)	1.8 (0.4, 8.7)	2.0 (0.4, 9.8)	2.3 (0.5, 11.0)	2.4 (0.5, 12.3)	2.1 (0.4, 9.9)	2.3 (0.4, 11.7)
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs2070673								
AT	0.7 (0.3, 1.7)	0.7 (0.3, 1.9)	0.6 (0.2, 1.7)	0.7 (0.2, 2.0)	0.6 (0.2, 1.9)	0.7 (0.2, 2.1)	0.6 (0.2, 1.7)	0.6 (0.2, 1.9)
AA	--	--	--	--	--	--	--	--

TT	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs6413432								
AT	1.5 (0.6, 3.6)	1.5 (0.6, 3.9)	1.4 (0.5, 3.7)	1.5 (0.6, 4.2)	1.9 (0.7, 5.0)	2.0 (0.7, 5.6)	1.7 (0.6, 4.4)	1.8 (0.6, 5.0)
TT	--	--	--	--	--	--	--	--
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
CYP3A4 rs2740574								
AG	--	--	--	--	--	--	--	--
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
CYP3A5 rs776746								
AA	9.3 (0.6, 153.5)	7.0 (0.4, 122.6)	11.0 (0.7, 181.0)	9.1 (0.5, 162.6)	13.3 (0.8, 220.5)	11.0 (0.6, 200.1)	12.6 (0.8, 209.1)	10.7 (0.6, 192.5)
AG	0.7 (0.2, 2.5)	0.8 (0.2, 2.8)	0.9 (0.2, 3.0)	0.9 (0.3, 3.4)	0.7 (0.2, 3.1)	0.7 (0.2, 3.4)	1.0 (0.3, 3.5)	1.1 (0.3, 4.2)
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
CYP4B1 rs2297810								
AG	1.4 (0.6, 3.0)	1.2 (0.5, 2.8)	1.4 (0.6, 3.3)	1.3 (0.5, 3.0)	1.3 (0.5, 3.4)	1.2 (0.5, 3.1)	1.7 (0.7, 4.1)	1.5 (0.6, 3.7)
AA	--	--	--	--	--	--	--	--

GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
GSTA1 rs3957356								
AA	2.1 (1.0, 4.8)	1.7 (0.7, 4.1)	2.3 (1.0, 5.3)	1.9 (0.8, 4.7)	3.1 (1.3, 7.7)	2.7 (1.0, 7.0)	2.7 (1.1, 6.5)	2.3 (0.9, 5.9)
AG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
GSTA1 rs3957356								
GG	0.6 (0.3, 1.5)	0.6 (0.3, 1.6)	0.8 (0.3, 1.9)	0.8 (0.3, 2.0)	0.8 (0.3, 2.2)	0.8 (0.3, 2.2)	0.7 (0.3, 1.9)	0.7 (0.3, 1.9)
AG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
GSTA1 rs3957356								
AG or GG	0.5 (0.2, 1.0)	0.6 (0.2, 1.4)	0.4 (0.2, 1.0)	0.5 (0.2, 1.3)	0.3 (0.1, 0.8)	0.4 (0.1, 1.0)	0.4 (0.2, 0.9)	0.4 (0.2, 1.1)
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
GSTM1 Undetermined								
0	2.2 (1.0, 4.9)	2.0 (0.9, 4.6)	2.5 (1.0, 6.2)	2.4 (1.0, 5.9)	2.4 (0.9, 6.2)	2.2 (0.8, 6.0)	2.7 (1.0, 6.9)	2.5 (0.9, 6.5)
1	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
2	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
GSTM3 rs7483								

AA	2.4 (0.8, 7.0)	2.8 (0.9, 8.6)	3.2 (1.1, 9.5)	3.9 (1.2, 12.3)	3.8 (1.1, 12.8)	4.5 (1.2, 16.0)	2.6 (0.8, 8.3)	3.3 (1.0, 11.2)
AG	1.0 (0.4, 2.3)	1.0 (0.4, 2.4)	1.2 (0.5, 3.0)	1.2 (0.5, 3.1)	1.6 (0.6, 4.3)	1.6 (0.6, 4.5)	1.0 (0.4, 2.5)	1.0 (0.4, 2.6)
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs1332018								
CC	1.1 (0.4, 2.9)	0.8 (0.3, 2.3)	1.1 (0.4, 2.9)	0.8 (0.3, 2.2)	0.9 (0.3, 2.7)	0.6 (0.2, 2.1)	1.4 (0.5, 3.9)	1.0 (0.3, 3.0)
AC	0.4 (0.2, 0.9)	0.3 (0.1, 0.7)	0.3 (0.1, 0.8)	0.2 (0.1, 0.7)	0.3 (0.1, 0.9)	0.3 (0.1, 0.7)	0.4 (0.1, 1.1)	0.3 (0.1, 0.9)
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs1799735								
TG	0.5 (0.2, 1.3)	0.5 (0.2, 1.4)	0.4 (0.1, 1.3)	0.5 (0.2, 1.4)	0.4 (0.1, 1.4)	0.4 (0.1, 1.6)	0.5 (0.2, 1.5)	0.6 (0.2, 1.7)
TT	--	--	--	--	--	--	--	--
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
GSTM3 rs7483								
AA	2.4 (0.9, 6.4)	2.8 (1.0, 7.9)	2.9 (1.1, 7.7)	3.5 (1.2, 10.0)	3.0 (1.0, 8.7)	3.5 (1.1, 10.7)	2.6 (0.9, 7.7)	3.3 (1.1, 10.3)
AG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								

GSTP1								
rs1695								
AG	0.8 (0.4, 1.8)	0.8 (0.4, 1.8)	0.8 (0.4, 1.9)	0.8 (0.4, 1.9)	0.8 (0.3, 1.8)	0.7 (0.3, 1.8)	0.8 (0.3, 1.8)	0.7 (0.3, 1.8)
GG	--	--	--	--	--	--	--	--
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs1138272								
AA	1.0 (0.3, 3.0)	1.0 (0.3, 3.3)	1.2 (0.4, 3.6)	1.2 (0.4, 4.0)	1.0 (0.3, 3.6)	1.1 (0.3, 4.0)	1.0 (0.3, 3.4)	1.0 (0.3, 3.8)
AG	0.3 (0.0, 2.2)	0.3 (0.0, 2.3)	0.3 (0.0, 2.6)	0.4 (0.0, 2.9)	--	--	0.4 (0.0, 2.9)	0.4 (0.0, 3.2)
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
GSTT1								
Undetermined								
0	1.0 (0.4, 2.8)	0.9 (0.3, 2.6)	0.9 (0.3, 2.8)	0.8 (0.3, 2.5)	0.5 (0.1, 2.4)	0.5 (0.1, 2.1)	0.8 (0.2, 2.7)	0.6 (0.2, 2.4)
1	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
2	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
GSTZ1								
rs7972								
AA	1.6 (0.2, 13.9)	2.6 (0.3, 25.4)	1.9 (0.2, 16.7)	2.8 (0.3, 28.7)	2.5 (0.3, 22.2)	4.0 (0.4, 41.2)	2.2 (0.3, 19.6)	3.4 (0.3, 37.3)
AG	1.1 (0.4, 3.0)	1.1 (0.4, 3.2)	1.3 (0.5, 3.7)	1.4 (0.5, 3.9)	1.7 (0.6, 4.9)	1.8 (0.6, 5.3)	1.5 (0.5, 4.3)	1.6 (0.5, 4.8)
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
rs1046428								

AG	0.7 (0.3, 1.6)	0.5 (0.2, 1.3)	0.6 (0.2, 1.4)	0.4 (0.2, 1.1)	0.4 (0.1, 1.3)	0.3 (0.1, 1.0)	0.5 (0.2, 1.4)	0.4 (0.1, 1.1)
AA	--	--	--	--	--	--	--	--
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
JAK2 V617F								
rs12340895								
CC	4.7 (1.7, 12.9)	5.3 (1.8, 15.7)	7.4 (2.5, 22.2)	8.7 (2.6, 28.9)	9.0 (2.7, 30.8)	10.2 (2.8, 37.9)	8.6 (2.7, 27.0)	10.8 (3.0, 38.2)
CG	1.4 (0.6, 3.3)	1.5 (0.6, 3.6)	2.2 (0.8, 5.8)	2.5 (0.9, 6.8)	2.5 (0.8, 7.7)	2.6 (0.8, 8.3)	2.1 (0.7, 6.1)	2.4 (0.8, 7.1)
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs12343867								
GG	3.1 (1.0, 9.6)	4.3 (1.3, 14.7)	4.8 (1.4, 16.4)	7.2 (1.9, 27.5)	5.4 (1.3, 21.6)	7.9 (1.8, 34.9)	5.6 (1.6, 19.9)	9.6 (2.3, 39.8)
AG	1.3 (0.5, 3.1)	1.3 (0.5, 3.3)	2.0 (0.7, 5.4)	2.2 (0.8, 6.2)	2.2 (0.7, 6.9)	2.3 (0.7, 7.5)	1.8 (0.6, 5.5)	2.1 (0.7, 6.5)
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
NAT2								
rs1041983								
AA	0.5 (0.1, 2.2)	0.5 (0.1, 2.1)	0.6 (0.1, 2.6)	0.6 (0.1, 2.5)	0.7 (0.2, 3.3)	0.7 (0.1, 3.1)	0.3 (0.0, 2.4)	0.3 (0.0, 2.2)
AG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent

GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs1495741								
AA	3.2 (1.3, 8.0)	3.0 (1.2, 7.6)	3.4 (1.2, 9.2)	3.1 (1.1, 8.7)	4.9 (1.4, 16.8)	4.6 (1.3, 16.1)	3.8 (1.3, 11.5)	3.5 (1.1, 10.8)
AG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs1799929								
AA	2.3 (1.1, 5.1)	2.3 (1.0, 5.2)	2.5 (1.1, 5.7)	2.4 (1.0, 5.7)	3.5 (1.4, 8.5)	3.4 (1.3, 8.4)	2.5 (1.1, 6.1)	2.4 (1.0, 6.1)
AG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
rs1799930								
AA	0.3 (0.0, 2.4)	0.3 (0.0, 2.2)	0.3 (0.0, 2.7)	0.3 (0.0, 2.6)	0.4 (0.1, 3.3)	0.4 (0.0, 3.1)	--	--
AG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs1801279								
AA	1.0 (., .)	1.0 (., .)	1.0 (., .)	1.0 (., .)	1.0 (., .)	1.0 (., .)	1.0 (., .)	1.0 (., .)
AG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent

Missing								
rs1801280								
AA	0.6 (0.3, 1.4)	0.6 (0.2, 1.4)	0.6 (0.3, 1.5)	0.6 (0.2, 1.5)	0.5 (0.2, 1.4)	0.5 (0.2, 1.4)	0.5 (0.2, 1.3)	0.4 (0.2, 1.2)
AG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
NAT2								
rs1041983								
GG	0.9 (0.4, 1.9)	1.0 (0.5, 2.1)	1.0 (0.5, 2.2)	1.1 (0.5, 2.4)	1.3 (0.5, 3.1)	1.4 (0.6, 3.3)	1.1 (0.5, 2.5)	1.2 (0.5, 2.8)
AG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs1495741								
GG	0.4 (0.1, 3.1)	0.4 (0.0, 3.0)	0.5 (0.1, 3.6)	0.5 (0.1, 3.7)	0.6 (0.1, 4.5)	0.6 (0.1, 4.6)	0.5 (0.1, 4.1)	0.5 (0.1, 4.3)
AG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs1799929								
GG	0.8 (0.3, 1.7)	0.8 (0.3, 1.7)	0.8 (0.3, 1.9)	0.8 (0.3, 1.9)	0.8 (0.3, 1.9)	0.8 (0.3, 2.0)	0.7 (0.3, 1.6)	0.7 (0.3, 1.7)
AG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent

AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
rs1799930								
GG	1.0 (0.5, 2.0)	1.1 (0.5, 2.3)	1.0 (0.5, 2.3)	1.1 (0.5, 2.6)	1.4 (0.6, 3.4)	1.5 (0.6, 3.7)	1.1 (0.5, 2.6)	1.3 (0.5, 3.0)
AG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs1801279								
GG	1.0 (., .)	1.0 (., .)	1.0 (., .)	1.0 (., .)	1.0 (., .)	1.0 (., .)	1.0 (., .)	1.0 (., .)
AG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs1801280								
GG	2.5 (1.2, 5.5)	2.6 (1.2, 5.8)	2.8 (1.2, 6.3)	2.8 (1.2, 6.5)	4.0 (1.7, 9.7)	4.0 (1.6, 9.9)	2.9 (1.2, 6.8)	2.9 (1.2, 7.1)
AG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
NAT2								
rs1041983								
AG or GG	2.0 (0.5, 8.7)	2.1 (0.5, 9.4)	1.7 (0.4, 7.5)	1.8 (0.4, 8.2)	1.4 (0.3, 6.1)	1.5 (0.3, 6.8)	3.1 (0.4, 24.0)	3.5 (0.5, 27.9)
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								

rs1495741								
AG or GG	0.3 (0.1, 0.8)	0.3 (0.1, 0.9)	0.3 (0.1, 0.8)	0.3 (0.1, 0.9)	0.2 (0.1, 0.7)	0.2 (0.1, 0.8)	0.3 (0.1, 0.8)	0.3 (0.1, 0.9)
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs1799929								
AG or GG	0.4 (0.2, 1.0)	0.4 (0.2, 1.0)	0.4 (0.2, 0.9)	0.4 (0.2, 1.0)	0.3 (0.1, 0.7)	0.3 (0.1, 0.7)	0.4 (0.2, 0.9)	0.4 (0.2, 1.0)
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
rs1799930								
AG or GG	3.2 (0.4, 24.4)	3.5 (0.4, 27.1)	2.9 (0.4, 22.0)	3.0 (0.4, 23.8)	2.3 (0.3, 17.9)	2.5 (0.3, 20.1)	--	--
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs1801279								
AG or GG	1.0 (., .)	1.0 (., .)	1.0 (., .)	1.0 (., .)	1.0 (., .)	1.0 (., .)	1.0 (., .)	1.0 (., .)
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs1801280								
AG or GG	1.6 (0.7, 3.8)	1.7 (0.7, 4.1)	1.6 (0.7, 4.0)	1.7 (0.7, 4.2)	1.9 (0.7, 5.4)	2.0 (0.7, 5.7)	2.2 (0.8, 6.0)	2.3 (0.8, 6.5)
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs854560								

AT	0.9 (0.4, 1.8)	0.9 (0.4, 1.9)	0.7 (0.3, 1.6)	0.7 (0.3, 1.7)	0.8 (0.3, 1.9)	0.8 (0.3, 2.0)	0.7 (0.3, 1.6)	0.7 (0.3, 1.6)
TT	0.2 (0.0, 1.7)	0.2 (0.0, 1.6)	0.2 (0.0, 1.8)	0.2 (0.0, 1.7)	0.3 (0.0, 2.3)	0.3 (0.0, 2.1)	0.2 (0.0, 1.9)	0.2 (0.0, 1.8)
AA	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
NQO1								
rs1131341								
AG	1.7 (0.5, 6.3)	1.5 (0.4, 5.7)	2.0 (0.6, 7.4)	1.7 (0.4, 6.6)	2.6 (0.7, 9.5)	2.3 (0.6, 8.8)	1.5 (0.3, 6.8)	1.2 (0.2, 5.6)
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
rs1800566								
AA	0.9 (0.1, 7.7)	1.3 (0.1, 11.2)	1.1 (0.1, 9.5)	1.6 (0.2, 14.4)	--	--	1.2 (0.1, 10.3)	2.0 (0.2, 17.9)
AG	1.7 (0.8, 3.6)	1.7 (0.8, 3.7)	1.9 (0.8, 4.3)	1.9 (0.8, 4.3)	1.8 (0.8, 4.4)	1.8 (0.7, 4.4)	1.7 (0.7, 4.0)	1.7 (0.7, 4.1)
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
CYP1A1								
rs1048943								
CT	1.7 (0.5, 6.0)	1.8 (0.5, 6.8)	1.9 (0.5, 7.1)	2.1 (0.5, 8.1)	2.4 (0.6, 8.7)	2.4 (0.6, 9.5)	1.4 (0.3, 6.6)	1.4 (0.3, 7.0)
Tp53								
rs1042522								
GG	-	-	4.1 (1.2,14.5)	5.4 (1.4, 20.8)	5.9 (1.6, 22.1)	8.0 (2.0, 33.0)	4.8 (1.3, 17.5)	6.8 (1.7, 27.7)
CG	-	-	1.0	1.0	1.4	1.3	1.1	1.0

			(0.4, 2.5)	(0.4, 2.3)	(0.5, 3.6)	(0.5, 3.4)	(0.4, 2.8)	(0.4, 2.6)
CC	-	-	Referent	Referent	Referent	Referent	Referent	Referent
T	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								
rs4646903								
CC	3.8 (0.4, 38.1)	4.8 (0.4, 53.6)	4.2 (0.4, 42.3)	5.1 (0.5, 57.5)	--	--	4.7 (0.5, 47.5)	6.8 (0.6, 80.3)
CT	2.0 (0.8, 4.7)	1.9 (0.8, 4.7)	1.6 (0.6, 4.3)	1.5 (0.6, 4.2)	1.6 (0.6, 4.7)	1.6 (0.5, 4.6)	1.5 (0.5, 4.3)	1.4 (0.5, 4.0)
TT	Referent	Referent	Referent	Referent	Referent	Referent	Referent	Referent
Missing								

1. All cases include PV, ET and PMF, confirmed cases by expert panel include PV, ET, and PMF, confirmed JAK2 V617F mutation with PV, ET or PMF by expert panel, confirmed by expert panel PV only
2. Adjusted (for age, sex and county) Odds Ratios (95% confidence intervals)

Table 3.8 Total number deleterious number and frequency

Number of Deleterious genes[1]	Number and frequency deleterious genes[1]		Frequency deleterious genes[1]	
	Cases	Controls	Cases (%)	Controls (%)
0	0	11	0	3.70
1	1	99	3.23	33.30
At least 2	30	182	96.78	63.00

1. AHR, CY1A2, CY1E2, GSTA1, GSTM1, GSTM3, GSTT1, NAT2, TP53
2. Odds ratio adjusted for age, sex and county

Chapter 4: The Risk of Myeloproliferative Neoplasms Due to the Joint Effects of Susceptible Genotypes and Distance of Current Residence from Facilities with Known Hazardous Emissions

Abstract

Introduction

The causes of a rare category of myeloproliferative neoplasms (MPN)—a cancer of the bone marrow with an excess of blood cells—are currently unknown. These classical MPNs (polycythemia vera (PV), essential thrombocythemia (ET) and primary myelofibrosis (PMF)) share the same *de novo* somatic JAK2 V617F point mutation, but are clinically presented differently. Our recent work identified associations of genetic variations that affect toxicity of xenobiotics with risk of MPNs.

Objectives

To investigate the joint effects of previously identified environmentally sensitive genotypes implicated in MPNs and distance of residence from facilities with suspected hazardous emissions on risk of PV-related outcomes.

Methods

We conducted a population-based case-control study among residents of three Pennsylvania counties where a cluster of PV was previously described. Subjects were born between 1921 and 1968 and resided in the three counties between 2000 and 2008. Cases were identified from the Pennsylvania cancer registry and the previous cluster investigation, with multiple case categorizations including all MPNs combined, JAK2 V617F, and PV. Controls were selected using random-digit dialling and were screened for eligibility. Blood samples were obtained from participants who consented to blood collection (31 cases and 292 controls). These samples were genotyped for a panel of environmentally sensitive genes. Results from gene-only analysis was used to focus current analysis on genotypes that had a 2-4 fold increase in risk of an MPN. Data was

analyzed using logistic regression models that controlled for the design variables (age, sex and county) and evaluated one genetic variant at a time.

Information on hazardous waste sites in the tri-county area was obtained directly from the US Environmental Protection Agency, CERCLIS database. To detect the potential for existence of gene-environment interaction, the main effects of genes and distance from hazardous wastes sites (measure of exposure) as a continuous variable were used in this analysis that controlled for design variables (age and sex). Logistic regression was used to estimate odds ratios (OR) and 95% confidence intervals (CI).

Results

Overall continuous distance to any one of the six hazardous waste/waste-coal plant adjusted for age and sex were not associated in MPN risk with adjusted OR=0.97 (95%CI: 0.93, 1.01). Adjusted ORs living within 10 miles of a cogeneration plant or hazardous waste site had point estimates range here of 1.3 to 1.4 but with wide (95%CI:0.5 to 4) including the null. We also examined nine genotypes that suggested a signal when we looked at each risk genotype by distance category, the effect was not consistent, greater in some of the genotypes but not for all, compared to looking at the distance only model in Table 4.2. For AHR, GSTA1, GSTM3, and GSTT1 and TP53, these results were oscillating back at forth around the null for distances less than or equal to 10 miles with overlapping confidence intervals. NAT2 and CYP1A2 provided a steady effect estimate in the same direction with from a 1 – 20 fold increase of risk of an MPN for both distance categories, although it was very imprecise.

Conclusions

Overall distance from one of six hazardous waste/waste-coal plant examined in this study was not related to MPN risk. The sources that were examined maybe incorrect as not all sources were included or were characterized with too much error to show an effect. It is also possible that the sensitive genotype associations were due to chance and there is no environmental influence. While the gene-only results did not confirm a gene-environment interaction effect, it did offer the opportunity to explore the joint effects of using distance as a proxy for exposure with respect to NAT2, GST, and CYP gene biological pathway and chemical exposures.

Introduction

Myeloproliferative disorders are now called Myeloproliferative neoplasms or (MPNs), according to the World Health Organization's (WHO) new classification of hematopoietic tumors. There are no known causes of MPNs (Seaman et al., 2009). These rare cancers are characterized by an overproduction of red blood cells, platelets, and white cells. MPNs include polycythemia vera (PV), essential thrombocythemia (ET), and primary myelofibrosis (PMF) (Campbell & Green, 2005). Patients with PV can have>25% above predicted red cell mass (Campbell & Green, 2005). The overproduction of platelets in the bone marrow is seen in ET cases, and scarring of the bone occurs in PMF cases (Tefferi et al., 2007).

MPNs were classified as a reportable cancer by the WHO in 2001, and they revised the diagnostic criterion in 2008. The National Cancer Institute estimated the incidence for the years 2001-2004 of PV at 2.8 per 100,000 persons per year (Kutti &Ridell, 2001). Incidence of 0.4 cases per 100,000 persons per year was estimated for PMF and 1.5 cases per 100,000 per year for ET (Kutti &Ridell, 2001).

In 2009, the Centers for Disease Control (CDC) identified a PV cluster in Northeast Pennsylvania. This was first reported cluster for any MPN. MPNs with an unknown etiology create the possibility that the presence of numerous potential environmental exposures in the cluster area, with no other known cause, could be associated with the cluster. With the etiology of these MPNs unknown, it creates the possibility that the presence of numerous potential environmental exposures in the cluster area, with no other known cause, could be associated with the cluster (Anderson et al.2012; Hoffman &Rondelle, 2007).

MPNs and the role of the JAK2 V617F mutation

In 2005 the presence of the JAK2 V617F somatic point mutation was identified with these classic MPNs. The acquired JAK2 V617F mutation is a protein that acts as an on-and-off switch regulating bone marrow activity and is suspected to be specifically characteristic of PV with clinical diagnosis of ET and PMF (Baxter, 2005). The tyrosine kinase complex is activated by the JAK2 V617F protein normally responsible for regulating blood cell production through molecular signalling in the bone marrow. As a result of this acquired mutation, the normal inhibition of growth is disrupted, resulting in increased blood cell production (Spivak et al., 2010). This JAK2 V617F mutation is not seen in any other cancers to date, but is present in nearly all PV patients (>95%) and about half of those with ET or PMF (Kralovics et al., 2007**;** Tefferi et al., 2007). These MPNs, PV, ET, and PMF commonly share the same JAK2 V617F somatic mutation, which could play a role in disease initiation or progression (Kralovics et al., 2007). The etiology of the JAK2 V617F mutation is currently unknown (Seaman et al., 2009).

In 2009, a germ-line haplotype—the 46/1 haplotype—was associated with PV patients who were JAK2 V617F, and is now considered to be the only risk factor for the JAK2 V617F mutation in PV patients (Jones et al., 2009;Andrikovics et al., 2010). In European populations, using genome-wide studies (GWAS), this 46/1 haplotype was found in 5-41% of those with JAK2 V617F MPNs (Jones et al., 2009; Olcaydu et al., 2009; Kilpivaara et al., 2009).

Familial clustering of PV, ET, and PMF

A family history of MPNs is associated with an increased risk of developing an MPN (Anderson et al., 2012, Landgren et al., 2008). In 2009, a large population-based study in Sweden showed a 5-7-fold increase in risk of developing an MPN for first-degree relatives of patients followed in 24,577 first degree relatives of 11,039 MPN patients who

were diagnosed with an MPN from 1958 to 2005 (Landgren et al., 2008). The role of inherited genetics is thought to influence both phenotype and susceptibility to MPNs (Andrikovics et al., 2010).

Environmental exposures and MPNs

After identification of the acquired JAK2 V617F mutations in MPN patients, the interest in the etiology of the disease has gained interest. Notably, there has been more research into possible links to environmental exposures (Seaman et al., 2009; Seaman et al,2010). However, there is limited evidence in the literature to suggest an association with MPNs and occupational exposures. Studies done by Kaplan, (1986), Terreros et al., (1997), and Mele et al. (1997) on industry occupations suggests benzene as a risk factor. Only Terreros, with nine cases of myeloproliferative syndrome, reported a significant association with assessment of benzene exposure specifically. Exposure to petroleum was found to be associated with an increased risk of PV, ET with an elevated Standard Mortality Rate (SMR) of petroleum refiner workers of 455, (95%CI: 120, 1164) when compared to the general population with a SMR of 88 (Kaplan, 1986). Quiroga et al (1981) reported an association, $p=0.002$, with exposure to petroleum. In this hospital based case control study, out of 51, cases, there were only 10 PV cases and four myeloproliferative disorders (Quiroga, 1981). However, Pasqualetti et al. (1991) reported a significant risk of hematological malignancies with exposure to aromatic hydrocarbons. However, no association was found with only myeloproliferative syndrome cases (n=44, ORs not reported). All of these studies are of limited size and did not use the molecular markers such as the JAK2 V617F mutation that have significantly improved the identifying of true cases of MPNs (Seaman et al., 2009).

In addition, in our study titled *The Risk of Myeloproliferative Neoplasms due to Exposure to Polycyclic Aromatic Hydrocarbons* (described elsewhere in Chapter 2), no relationships between MPNs and residential history, diet, lifestyle behaviors, or presumed exposure to aromatic and heterocyclic amines was found (Gross-Davis, Chapter 2).

Susceptible genotypes associated with MPNs and JAK2 V617F mutation

Although functional Single Nucleotide Polymorphisms (SNPs) that may modify the biological dose of a chemical mutagen are being studied for a wide range of cancers, no studies have investigated any combination of these genotypes and the risk of MPNs.

The CYP1A2, NAT2, GSTM3 and GSTA1 SNPs came directly from Aim 2 of this dissertation where associations were detected with a 2-5-fold increase of risk of developing an MPN. An additional five SNPs, AHR, GSTA1, GSTM3, and GSTT1 and TP53, were included because they also met the biological pathway we are exploring, and our results from Aim2 suggested an association with MPNs. These SNPs are associated with the metabolism/detoxification of benzene, and were included based on recent publications surround benzene and MPNs (Schnatter et al., 2012).

The Cytochrome P450 superfamily (CYP), Polymorphic N-acetyltransferase (NAT2), and the Glutathione S-transferase (GST) polymorphisms have been identified in the literature and linked to increased cancer risk and continue to be included in research efforts. With *The Human Genome Project* at its completion, new methods continue to be published to conduct these gene-environment interaction studies. The candidate-gene approach, using a biological pathway hypothesis as its framework, is one of these methods used to examine gene-environment interactions (Hunter et al., 2005). The

National Institute of Environmental Health Sciences Environmental Genome Project (NIEHS EGP) has built a foundation to explore environmentally sensitive genes (ESG) and to categorize them in the human genome. Included in the NIEHS EGP list are a total of 648 genes in the following categories: gene expression; cell structure; cell cycle; metabolism; DNA repair; cell division; homeostasis; and cell signalling.

Polymorphisms affecting xenobiotic metabolizing enzymes have been used as markers for susceptibility (Their, 2003). These functional SNPs described below will encode enzymes for the biotransformation of chemicals such as PAHs and benzene. They can be genotoxic for cells or cause toxicant-induced damage to cellular DNA (Their, 2003). Observed hematologic effects with acute and chronic exposures have been well studied to show their direct association of hematologic changes in humans as well as animals (Galbraith et al., 2010; Smith, 2010).

Susceptible genotypes associated with MPNs, as reported by Gross-Davis (2013)

Cases and controls were similar in all demographic variables, as with the previous analysis of this population that investigated the role of genotypes that modify the toxicity of chemical mutagens in the risk for MPNs. As we reported, the prevalence of the susceptible genotypes was in agreement with the reported frequency in the literature (see Table 3.1 of MPNs Gross-Davis, Chapter 3). Our analysis of the main effects 14 environmentally sensitive genes, the presence of CYP1A2 and *GSTA1*, *GSTM3*, *NAT2* slow acetylator genotype, and TP53 (point estimates of adjusted (aORs) 2.7 to 4.6, with 95% C.I.s that excluded 1.0), variants were associated with an increased risk for all MPNs . All but one case harbored at least one deleterious SNP 30/31 (96.7%). out of the initial four SNPs, and all 31 cases had at least of the nine SNPs; AHR, CYP1E2, GSTM1, Tp53,

GSTT1, GSTM3, CYP1A2, NAT2, and GSTA1. Because our analysis found a 2 to 5 fold increase in developing an MPN associated with it prompted us to further examine the interaction hypothesis with respect to NAT2, GST, and CYP, AHR the gene variants with distances to industrial facilities.

Demographics

Tables 3.4 and 3.5 show the study demographics by case and controls status. 31 cases and 273 controls were interviewed and consented to genotyping and storage of bio-sample. Our cases were diagnosed with one of the following MPNs: PV 27 (90%), 2 (7%) with ET and 1 (>3%) PMF. Genotype data were available for 57% of the total cases and 62% of controls from the larger study described in Chapter 2 of this dissertation (Gross-Davis, 2013).

Cases were older (median age=71 vs. 63yrs) (OR=2.9, 95% CI: 1.6, 5.2) and more likely to be male (55% vs. 45%) (OR=1.4, 95%0. 8, 2.5) compared to controls but otherwise demographically similar (Tables 3.4 and Table 3.5). The study population was all US born and overwhelmingly Caucasian with no cases and few (2%) controls were of Jewish ancestry. Two-thirds of the study population was currently married and cases were more likely to be retired (65% versus 47% of controls).

No relationships were found between MPNs and residential history, lifestyle behaviors, and diet, with presumed exposure to aromatic and heterocyclic amines. Participation rates of cases were poor but the results do rule out strong associations between environmental PAHs and MPNs, with the null result for exposure to cigarette smoke (rich in PAH), providing the most convincing evidence.

Industrial sources in the Carbon, Luzerne, and Schuylkill Counties with potential hazardous emissions

In 2008, the joint efforts by the Pennsylvania Department of Health (PADOH) and the Agency for Toxic Substances and Registry (ATSDR) of the CDC confirmed a cancer cluster in a tri-county region of Northeast Pennsylvania. A cluster of cases was identified where the incidence of polycythemia vera is 4.3 times ($P < 0.001$) that of the rest of the study area (Seaman et al., 2009). The area of the cluster of cases in the 2009 report identified numerous sources of hazardous material including U.S. Environmental Protection Agency (EPA) Superfund sites and waste-coal power plants or co-generation plants (Seaman et al., 2009).

Several hazardous material exposure sources were found in or near the high-rate areas of PV. These plants began operations in the early 1990s, and their emissions are characterized by fine particulate matter, including complex hydrocarbons and various heavy metals like cadmium (Lui et al., 2008). There are U.S. EPA Superfund sites contained within the cluster area with potential offsite migration in the past (Seaman et al., 2009). These sites are described in more detail below.

Hazardous waste sites/waste-coal power plants

The Northeastern Power Company operates in McAdoo, PA. This is a 50-megawatt cogeneration station. Known as NEPCO, the plant is located in the eastern, anthracite coal-mining region of Pennsylvania. It uses a circulating fluidized bed boiler to combust anthracite coal mining refuse (culm). Steam from the plant is used in a 20-acre greenhouse specializing in flowers.

Wheelabrator Frackville Energy Company located in Frackville, PA. This 42-megawatt cogeneration station is located in eastern Pennsylvania. Steam from the plant is used to heat a nearby state correctional facility. The plant is located on the site of an abandoned coal mine and burns anthracite coal mining refuse in a circulating fluidized bed boiler.

WPS Westwood Generation LLC is located in Schuylkill County. WPS Westwood Generation, LLC operates in the Westwood Generating Station located in Tremont, Pennsylvania. The 30-megawatt power plant uses a Circulating Fluidized Bed boiler that converts waste coal into low-cost power. Steam from the plant drives a 40,000 horsepower generator that produces enough energy to serve 45,000 homes.

Tonolli Corporation is located along Rte. 54 in Nesquehoning, Carbon county Pennsylvania. The Tonolli site operated between 1974 and 1985 as a lead-acid battery recycling facility and a secondary lead smelter. The contaminants found at the site included lead, arsenic, cadmium, and chromium from the founder smelter. These were also found in monitoring wells and the soil, and later migrated to Nesquehoning creek. Cadmium toxicity is suspected to be genotoxic and is the main environmental exposure considered in this study. This site was formally added to the National Priorities List (NPL) in October 1989.

Eastern Diversified Metals (EDM) is a former recycling facility located in Rush Township, Schuylkill County. This 25-acre site disposed of approximately 350 million pounds of waste insulation material in an open pile, called “fluff”. The fluff pile at this site occupied a parcel of land approximately 1,500 feet long, up to 60 feet high and 250 feet wide. The threats at the site include polychlorinated biphenyls and dioxin compounds, along with other volatile organic compounds that may or may not be site-

related contaminates. These compounds contaminated the ground water that serves 1600 residents living within a one-mile radius.

McAdoo Associates is a site of interest to the residents of Schuylkill County due to past contamination concerns, located in the Borough of McAdoo. This site is currently still under remediation. Base Neutral Acids, Dioxins/Dibenzofurans, Inorganics, Metals, PAH, PCBs, Pesticides, and VOCs were found on site soil and groundwater monitoring wells. Work on McAdoo is not yet complete, as one operable unit continues to create groundwater contamination.

The objective of this study is to explore the identified susceptible genotypes and exposure to hazardous waste facilities in the risk for myeloproliferative neoplasms.

Methods

The study design and data collection were described elsewhere (Chapters 2 & 3) and are only briefly summarized below.

Study design

An unmatched case-control study design was used in this analysis. The study area was located in Northeast Pennsylvania and comprised a tri-county area (Carbon, Luzern, and Schuylkill Counties). Cases were individuals with diagnoses between 2001 and 2010 of PV, ET, PMF, with and without JAK2 V617F. Controls were selected using random digit dialling along with age, county, and the amount of time they had resided there. Both cases and controls had their DNA checked for JAK2 V617F mutation and a group of genes we were interested in AHR, CY1A2, CY1E2, GSTA1, GSTM1, GSTM3, GSTT1, NAT2, and TP53.

Participants

An estimated 500,000 people in 2007 resided in the tri-county region. We restricted the study population by age to eliminate younger individuals not at risk for an MPN. Since this study focuses on environmental exposures in the tri-county area, all cases and controls were required to reside in the tri-county area for between 2000 and 2008.

Case definition

Cases must have met the clinical criteria for an MPN (PV, ET or PMF), and received a diagnosis between the years 2001 and 2010. They also had to be born between 1/1/ 1921 and 12/31/1968 to be included. The following case categorization for eligible cases was used in this analysis: all MPNs cases that provided a blood sample for genotyping (see Table 2.2). Cases were identified from two sources. First, from the Pennsylvania Cancer Registry (PCR) using the International Classification of Diseases codes (ICD-O) for MPNs (codes M-9950/3, M-9962/3, M-9961/3, and M-9931/3). Second from a PV cluster investigation in 2009 (conducted by the Agency for Toxics Disease Registry (ATSDR)). Cases not confirmed in the ATSDR study were reviewed by either the PA DOH expert panel or by a concurrent University of Pittsburgh expert panel funded by PA DOH described in Chapter 2.

Controls selection

Controls were selected from the source population in the tri-county area. Selection was based on eligibility screening questions using random digit dialling. Residents born from 1921-1968 and living in these one of the three counties were eligible. A stratified random sample of controls was selected from the study area and was comparable to the population distribution described in the U.S. Census American Community Survey, 2008.

Data collection

We assessed lifetime residential, smoking, and dietary history by telephone interview. The two-phase consent required completion of the phone interview and a blood draw. A $25 gift card was offered as an incentive for each phase (Gross-Davis, Chapter 3, 2013). A full description of the consent procedure is described in *the role of genotypes that modify the toxicity of chemical mutagens in the risk for myeloproliferative neoplasms* study (Gross-Davis, Chapter 3, 2013).

A one-time blood draw of peripheral venipuncture of 25-30ml of blood was required to get ample blood samples for JAK2 V617F mutation testing, gene susceptibility testing, storage, and possible future testing for biomarkers linked to MPN by other studies ongoing in the tri-county area (PV Partners studies). Geisinger Health Systems collected the blood samples, and both Mt. Sinai and Geisinger Health Systems completed JAK2 V617Ftesting. Columbia University Laboratory extracted the DNA and stored the samples. A full description of the genotyping procedure used by Columbia University can be found in the original manuscript (Gross-Davis, Chapter 3, 2013).

Distance to hazardous waste/waste-coal sites as a proxy for exposure calculation

In order to test for joint effects of distance and susceptible genotype, proximity metric was employed. This proximity metric uses distance as the exposure measure and is based on the assumption that people living closer to these hazardous waste sites and waste-coal plants with suspected emissions have higher exposures than those living farther away. Each hazardous waste sites and waste-coal plants was analyzed separately

and stratified by two categories: <= 10 miles from the hazardous waste sites and waste-coal plants, and > 10 miles from each hazardous waste sites and waste-coal plants. All distance calculations were geocoded using SAS 9.2 mapping function (GeoCode) for automated matching. We used the residents' current addresses obtained during the original study to get latitude and longitude for each location. The first attempt to convert these addresses to a latitude and longitude was fairly successful. For controls, we had an 84.6% match of addresses to a latitude and longitude in the three counties, with the remaining 15.4 % matching at least for the zip code. Similarly, our first attempt for addresses of cases, matched a latitude and longitude in the 3 county areas for 78.2% of the total sample with the remaining (21.8%) matching at least for zip code. For cases that we could not obtain a latitude or longitude from SAS, we address-matched eight cases using Google Maps/Google Earth, found two others by using the US Postal Service website, and obtained the remaining two using the search engine Google. For most addresses that did not match, the house address was incorrect although all zip codes were successfully matched 100% on the first pass. Where the address obtained from the interview did not produce a latitude and longitude, we used the halfway point on street names to calculate distances to hazardous waste sites. The latitude and longitude of the hazardous waste sites were inputted directly into SAS. The CDC (via personal communication) gave us the waste-coal power plant latitude and longitude from an environmental assessment report that is going through peer review at the CDC. The hazardous waste site latitude and longitudes were obtained directly from the Comprehensive Environmental Response, Compensation, and Liability Information System (CERCLIS) database, maintained exclusively by the US Environmental Protection Agency.

Data analysis

Descriptive analysis was conducted for 31 cases and 273 controls on the characteristics of the study population. SAS version 9.2 logistic regression was used to estimate adjusted odds ratios (aOR) and associated 95% confidence intervals (CI) adjusted for design variables only, case and sex. To detect the potential for existence of gene-environment interaction, the main effects of genes and distance from hazardous wastes sites and waste-coal power plants (measure of exposure) were used. The measure of distance was analyzed by a continuous variable for marginal effects of distance. To test for joint effects of susceptible genotype and distance to hazardous wastes site or a waste-coal power plants distance was categorized as near (<= 10 miles) and far (>10 miles). We then stratified these results by each genotype to explore interaction without an interaction term, because our sample size was insufficient to accurately estimate cross-product interaction term in logistic regression. In addition, analysis of the number of deleterious SNPs as a continuous variable was stratified by distance using as near (<= 10 miles) and far (>10 miles), using logistic regression to estimate crude and adjusted OR s with 95% confidence intervals.

Results

Overall the distance from one of six hazardous waste/waste-coal power plants examined in this study was not related to MPN risk. Approximately one quarter of the cases lived within 10 miles of Tonolli Corporation, McAdoo Associates or Eastern Diversified Metals, Northeastern Power Company, Westwood generation, and Wheelabrator Frackville Energy facilities (Table 4.2). There is no suggestion in the data that an effect was missed using a 10-mile cut off as the near proximity as the mean distance for our cases and controls to each site did not dip below 17 .7 miles, with the 25th percentile for all cases between 8.8 miles and 23.4 miles away, and 11.1 to 26.8 miles away for controls

(Table 4.3). When all facilities were including in the analysis for living <=10 miles to each site was analyzed, there was not much change to the effect estimate, as so few cases live near any of these hazardous waste site or waste coal power plants (results not shown).

We also examined the effects of residential proximity to three waste-coal plants Northeastern Power Company, Westwood Generation, and Wheelabrator Frackville Energy and three hazardous waste sites, Tonolli Corporation, McAdoo Associates and Eastern Diversified Metals on MPN risk in a three county area (Table 4.2). There was a slight effect detected for all sites except Westwood generation, with no cases living within 10 miles from the site and no effect was observed for Wheelabrator Frackville Energy, (aOR=0.96, 95% CI:0.2, 3.7). Overall distance to any one of the six hazardous waste/waste-coal plants adjusted for age and sex were generally unimpressive.

All three hazardous waste sites showed similar effects, (EDM aOR=1.3, 95% CI: 0.5, 3, 3; McAdoo Associates with aOR= 1.5, 95%, CI: 0.6, 3.8 and Tonolli Corporation with an aOR= 1.4, 95%CI: 0.5, 3.9), although none reached a magnitude to elicit concern and were imprecise. We observed the strongest effect estimate from people living less than or equal 10 miles from the Northeastern Power Company, with an aOR=1.6, 95% CI: 0.64, 4.0).

Susceptible genotype stratified by distance

We explored nine specific genotypes in this study. Data describing these genotypes stratified by distance are presented in Table 4.3. When we looked at each risk genotype by distance category, the effect was not consistent, thus greater in some of the genotypes but not for all, compared to looking at the distance only model in Table 4.2. For AHR,

GSTA1, GSTM3, and GSTT1 and TP53, these results were not consistent, oscillating back at forth around the null for distances less than or equal to 10 miles with overlapping confidence intervals. The results were not the same for greater than 10 miles for all these SNPs. These dramatic swings in point estimates of the effect are due to small differences is the number of cases and controls living less than or equal to 10 miles from any particular hazardous waste/waste-coal plants, as shown in Table 4.1.

CYP1A2

CYP1A2 provided a steady effect estimate in the same direction with from a 2 – 20 fold increase of risk of an MPN for both distance categories, although it was very imprecise. For resident living <=10 miles of EDM, we also found close to 20 fold increase aOR=21 (95% CI: 0.6, 710) for CYP1A2 and for greater than 10 miles from Tonolli Corporation an aOR of 3.0 (95% CI: 0.8, 11). Specifically for residents harboring the CYP1A1 genotype and living > 10 miles from Tonolli Corporation, aOR=4.9 (95% CI: 1.4, 17.0), adjusted for age and county was detected however, the model failed to converge for the<= 10 miles category. Residents living near McAdoo, saw a similar effect size aOR=19 (95% CI: 0.6, 60) to EDM.

CYP1E2

Tonolli Corporation as well as Northeastern Power Company both showed an 8-fold increase for those subject who have the as risk allele for CY1E2 aOR=7.8 (95% CI: 0.7,80) and aOR=8.2 (95% CI: 0.3, 189) adjusted for age, and sex respectfully.

GSTM1, GSTA1, GSTM3

GSTM1 showed consistently a 2-3-fold increase of risk of an MPN for both near and far distance. We observe an adjusted OR of 2.9 (95% CI: 0.4,19.7) for residents living within

10 miles of the McAdoo Associates site, as well as an adjusted OR of 1.8 (95% CI:0.7, 4.7) for residents living greater than 10 miles from the McAdoo Associate site. The trend was similar for all hazardous waste sites and waste-coal power plants for GSTM1 null genotypes. For GSTM3 at risk genotypes, an effects estimate of 1 to 1.9 was observed only in residents living greater than 10 miles from the Tonolli Corporation site.

NAT2

Similarly for NAT2 slow acetylator subject living > 10 miles from Westwood Generation waste-coal power plant, we found an aOR=3.0 (95% CI: 1.1, 7.7) adjusted for age and sex, but the model also failed for residents living less than or equal to 10 miles from this plant.

Discussion

While the gene only results did not confirm a gene-environment interaction effect, it did offer the opportunity to explore the joint effects of using distance as a proxy for exposure with respect to NAT2, GST, and CYP gene and additional AHR, TP53 biological pathway and chemical exposures.

Although our results for genotypes associated with modifying benzene exposure is consistent with the suggestion by the Schnatter, et al., results for MDS cumulative exposure to benzene OR= 4.33 (95% CI: 1.31, 14.3) and 30 MPD cases (OR=1.79, 95% CI: 0.68, 4.74) (Schnatter, et al., 2012), the cases from both studies are diagnosed with different ICD-O coded diseases. This pooled analysis might have been more useful to our study, if when cases (medical records) were reviewed by the hematologists in the study, that the WHO 2008 classifications were also used for additional case

categorizations. The misclassification of this classis group of MPNs, still provided potential bias in their analysis (Miller, et al., 2010).

An expected limitation of using residential proximity (distance) for the estimated exposure is the relevant time frame this represents. In addition, proximity to hazardous waste/industrial sites could indicate multiple chemicals not accounted for in this analysis. All these sites do have potential emissions/releases of chemicals with potential mutagenic and/or genotoxic effects.

A potentially more serious problem was the small numbers of "highly exposed" subjects in some of the analysis, especially when considering specific industries. The sample size was maximized and distance was used as a continuous measure as opposed to exposed and unexposed. This assumption was critical for our focus which included susceptible genotypes where are all cases are considered “exposed” to some amount, but none to zero exposure to these ubiquitous chemicals. This assumption, however might have affected usefulness of proximity to these facilities as a measure of exposure.

Residential proximity to a hazardous waste site or an active waste-coal power plant does not, alone, indicate exposure, and acknowledging that distance can only weakly approximate dose. Although we did have a mechanism for validating actual distances of current addresses using the SAS mapping function, we did not validate actual individual exposure since no individual pollutants were identified or measured for this analysis. Nonetheless, there was the suggestion of a gradient with distance from reported industries. We did not analyze the duration of residence used in the calculation, as our previous findings concerning residence time were not useful. Thus, it is possible that the number of years living in a particular residence could lead to higher exposure than our

basic model. We did not attempt to aggregate durations of exposures to multiple industries. Instead, we evaluated the distance of residence near the each hazardous waste site individually.

Our response rate from MPN cases was small and the request for cases to consent to a bio-sample could have contributed to this. This may be a concern for bias. However, our randomized selection of controls for the tri-county area attempted to minimize any concern for bias. We had a 1 to 9 ratio of cases and controls to gain a representative comparison population and distribution of genotypes among controls is as can be expected in the source population with no significant demographic differences observed between case and controls that consented to genotyping and those who did not.

The latency period between possible exposures related to the development of an MPN is currently unknown, which complicates identification of potential risk factors when the timeframe of effective exposure is not clearly defined. So using current residence does attempt to capture exposure during an unknown window and was used as a simple approach to investigate these joint effects.

A strength of the study was the high percentage of cases and controls that were successfully geocoded on the first attempt. Errors were typically due to spelling or missing house numbers but there was also perfect matching for zip code for cases and controls. This study was also able to capitalize on the ability to test for JAK2 V617F mutation molecular testing to reduce case misclassification that is a concern with MPN research.

Our study did not showed any consistent relationship with MPNs and joint effects of proximity to known hazardous waste site and waste-coal power plants and environmentally sensitive genotypes, and the results we did observe were very imprecise.

List of References

Anderson, L. A., Duncombe, A. S., Hughes, M., Mills, M. E., Wilson, J. C., & McMullin, M. F. (2012). Environmental, lifestyle, and familial/ethnic factors associated with myeloproliferative neoplasms. *Am J Hematol, 87*(2), 175-182.

Andrikovics, H., Nahajevszky, S., Koszarska, M., Meggyesi, N., Bors, A., Halm, G., Tordai, A. (2010). JAK2 46/1 haplotype analysis in myeloproliferative neoplasms and acute myeloid leukemia. *Leukemia, 24*(10), 1809-1813.

Campbell, P. J., & Green, A. R. (2005). Management of polycythemia vera and essential thrombocythemia. *Hematology Am Soc Hematol Educ Program*, 201-208

Clapp, R. W., Jacobs, M. M., & Loechler, E. L. (2008). Environmental and occupational causes of cancer: new evidence 2005-2007. *Rev Environ Health, 23*(1), 1-37.

Dong, Linda M., Potter, John D., White, Emily, Ulrich, Cornelia M., Cardon, Lon R., & Peters, Ulrike. (2008). Genetic Susceptibility to Cancer. *JAMA: The Journal of the American Medical Association, 299*(20), 2423-2436.

Galbraith, D., Gross, S. A., & Paustenbach, D. (2010). Benzene and human health: A historical review and appraisal of associations with various diseases. *Crit Rev Toxicol, 40 Suppl 2*, 1-46.

Garcia-Closas, M., et al. (2013). "Common Genetic Polymorphisms Modify the Effect of Smoking on Absolute Risk of Bladder Cancer." Cancer Research 73(7): 2211-2220.

Glass DC, Gray CN, Jolley DJ, et al. Leukemia risk associated with low-level benzene exposure. Epidemiology 2003. 14(5):569–577.

Green J, Banks E, Berrington A, Darby S, Deo H, Newton R. N- □acetyltransferase 2 and bladder cancer: an overview and consideration of the evidence for gene-environment interaction. Br J Cancer 2000; 83:412–7.

Gross-Davis, CA. (2013)Environmental Etiology of Polycythemia Vera, Essential Thrombocythemia and Primary Myelofibrosis: A Case-Control Study in Northeast Pennsylvania disseration, Drexel Univeristy, School of Public Health

Hein DW, Doll MA, Rustan TD, Gray K, Feng Y, Ferguson RJ, Grant DM. Metabolic activation and deactivation of arylamine carcinogens by recombinant human NAT1 and polymorphic NAT2 acetyltrans- ferases. Carcinogenesis 1993;14:1633–8

Hoffman R, Rondelli D. Biology and treatment of primary myelofibrosis. Hematology Am Soc Hematol Educ Program 2007;2007:346-354.

Hung, R. J., Boffetta, P., Brockmoller, J., Butkiewicz, D., Cascorbi, I., Clapper, M. L., et al. (2003). CYP1A1 and GSTM1 genetic polymorphisms and lung cancer risk in Caucasian non-smokers: a pooled analysis. *Carcinogenesis, 24*(5), 875-882.

Hunter, D. J. (2005). Gene–environment interactions in human diseases. *Nature Reviews Genetics*, *6*(4), 287-298.

Jones, Amy V., Chase, Andrew, Silver, Richard T., Oscier, David, Zoi, Katerina, Wang, Y. Lynn, . . . Cross, Nicholas C. P. (2009). JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nat Genet, 41*(4), 446-449.

Kaplan, S. D. (1986). Update of a mortality study of workers in petroleum refineries. *J Occup Med, 28*(7), 514-516.

Kilpivaara, Outi, Mukherjee, Semanti, Schram, Alison M., Wadleigh, Martha, Mullally, Ann, Ebert, Benjamin L., Levine, Ross L. (2009). A germline JAK2 SNP is associated with predisposition to the development of JAK2V617F-positive myeloproliferative neoplasms. *Nat Genet, 41*(4), 455-459.

Kralovics R, Passamonti F, Buser AS, Teo S, Tiedt R, Passweg JR, Tichelli A, Cazzola M, Skoda RC. A gain-of-function mutation of JAK2 in myeloproliferative disorders. N. Engl. J. Med. 2005;352:1779–1790.

Landgren, O., Kristinsson, S. Y., Goldin, L. R., Caporaso, N. E., Blimark, C., Mellqvist, U. H., Turesson, I. (2008). Risk of plasma cell and lymphoproliferative disorders among 14621 first-degree relatives of 4458 patients with monoclonal gammopathy of undetermined significance in Sweden. *Blood, 114*(4), 791-795.

Liu, G., Niu, Z., Van Niekerk, D., Xue, J., & Zheng, L. (2008). Polycyclic aromatic hydrocarbons (PAHs) from coal combustion: emissions, analysis, and toxicology. In Reviews of environmental contamination and toxicology (pp. 1-28). Springer New York.

Malcovati, L., & Cazzola, M. (2008). Myelodysplastic/myeloproliferative disorders. Haematologica, 93(1), 4-6.

Mele A, Visani G, Pulsoni A, et al. Risk factors for essential thrombocythe- mia—A case–

control study. Cancer 1996;77:2157–2161.

Miller, B. G., Fransman, W., Heederik, D., Hurley, J. F., Kromhout, H., & Fitzsimons, E. (2010). A review of the data quality and comparability of case–control studies of low-level exposure to benzene in the petroleum industry. *International archives of occupational and environmental health*, *83*(1), 69-76.

Moliterno, A. R., Williams, D. M., Rogers, O., Isaacs, M. A., & Spivak, J. L. (2008). Phenotypic variability within the JAK2 V617F-positive MPD: roles of progenitor cell and neutrophil allele burdens. *Exp Hematol, 36*(11), 1480-1486.

Olcaydu, Damla, Harutyunyan, Ashot, Jager, Roland, Berg, Tiina, Gisslinger, Bettina, Pabinger, Ingrid, . . . Kralovics, Robert. (2009). A common JAK2 haplotype confers susceptibility to myeloproliferative neoplasms. *Nat Genet, 41*(4), 450-454.

Pasqualetti, P., Casale, R., Colantonio, D., & Collacciani, A. (1991). Occupational risk for hematological malignancies. *Am J Hematol, 38*(2), 147-149.

Quiroga Micheo E, Calcagno EJ, Calabria SI, et al. Retrospective epidemio- logical study of hemopoietic system neoplasms in Argentina. Medicina 1981;41:187–200.

Rushton L, Romaniuk H. A case-control study to investigate the risk of leukaemia associated with exposure to benzene in petroleum marketing and distribution workers in the United Kingdom. Occup Environ Med 1997. 54(3):152–166.

Schnatter AR, Armstrong TW, Nicolich MJ, et al. Lymphohaematopoietic malignancies and quantitative estimates of exposure to benzene in Canadian petroleum distribution workers. Occup Environ Med 1996. 53(11):773–781.

Schnatter, Robert Glass, Deborah C. Tang, Gong, Irons, Richard D. and Rushton,Lesley. (2012) Myelodysplastic (2012). Syndrome and Benzene Exposure Among Petroleum Workers: An International Pooled Analysis. JNCI J Natl Cancer Inst.

Seaman, V., Dearwent, S. M., Gable, D., Lewis, B., Metcalf, S., Orloff, K., . . . Cole, H. (2010). A multidisciplinary investigation of a polycythemia vera cancer cluster of unknown origin. *Int J Environ Res Public Health, 7*(3), 1139-1152.

Seaman, V., Jumaan, A., Yanni, E., Lewis, B., Neyer, J., Roda, P., . . . Hoffman, R. (2009). Use of molecular testing to identify a cluster of patients with polycythemia vera in eastern Pennsylvania. *Cancer Epidemiol Biomarkers Prev, 18*(2), 534-540.

Smith, Martyn T. (2010). Advances in Understanding Benzene Health Effects and Susceptibility. *Annual Review of Public Health, 31*(1), 133-148.

Spivak JL. (2010) Narrative review: Thrombocytosis, polycythemia vera, and JAK2 mutations:The phenotypic mimicry of chronic myelofibrosis. Ann Intern Med. 2010 Mar 2;152(5):300-6. Review.

Swerdllow, S. H., Campo, E., & Harris, N. L. (2008). *WHO classification of tumours of haematopoietic and lymphoid tissues*. France: IARC Press, 2008.

Tefferi, A., Thiele, J., Orazi, A., Kvasnicka, H. M., Barbui, T., Hanson, C. A., . . . Vardiman, J. W. (2007). Proposals and rationale for revision of the World Health Organization diagnostic criteria for polycythemia vera, essential thrombocythemia, and primary myelofibrosis: recommendations from an ad hoc international expert panel. *Blood, 110*(4), 1092-1097.

Tefferi, A., & Vainchenker, W. (2011). Myeloproliferative neoplasms: molecular pathophysiology, essential clinical understanding, and treatment strategies. *J Clin Oncol, 29*(5), 573-582.

Terreros MC, Apezteguia M, Slavutsky IR, Guimarey LM. Exposure to occupational and environmental risk factors in hematologic disorders. Neoplasia 1997;14:133–136.

Thier, Ricarda, et al. "Markers of genetic susceptibility in human environmental hygiene and toxicology: the role of selected CYP, NAT and GST genes." *International journal of hygiene and environmental health* 206.3 (2003): 149-171

Vijayakrishnan, J., & Houlston, R. S. (2010). Candidate gene association studies and risk of childhood acute lymphoblastic leukemia: a systematic review and meta-analysis. *Haematologica, 95*(8), 1405-1414.

Wargovich, M. J., & Cunningham, J. E. (2003). Diet, individual responsiveness and cancer prevention. *J Nutr, 133*(7 Suppl), 2400S-2403S.

Table 4.1 Distance rounded to the nearest mile for cases and controls to each hazardous/waste-coal site

	EDM		McAdoo		Northeastern Power Company		Tonolli Corporation		Westwood Generation		Wheelabrator Frackville Energy	
	Cases n=25	Control n=292	Cases n=25	Control n=292	Cases n=25	Control n=292	Cases n=25	Control n=292	Cases n=25	Control n=292	Cases n=25	Control n=292
Mean	17.9	20.7	16.8	19.35	17.7	20.3	18.4	21	36.2	37.9	21.1	23.7
Minimum	2.4	1.1	3.0	1.5	2.2	4.9	5.5	1.2	10.5	4.8	2.6	2.8
25th percentile	10.4	12.7	8.8	11.1	9.8	11.6	9.7	12.4	24.3	26.8	10.7	12.9
Median	16.8	18.6	18.8	18.9	15.2	18.1	19.3	21.6	38.4	39.4	20.1	21.8
75th percentile	26.3	30.2	30.2	27.1	24.8	29.4	26.0	28.9	49.1	53	31	35.9
Maximum	36.5	40.6	40.6	37.4	36.2	40.0	33.9	40.4	60.2	63.2	43.1	46.6

Table 4.2 Number and frequency of cases and controls with residence near (<= 10 miles) and far (> 10 miles) to Hazardous waste site / waste-coal power plant

Hazardous waste site / waste-coal power plant	Number and frequency (%) Residence < =10 miles		Number and frequency (%) Residence >10 miles		Over continuous distance measured in miles		Residence < =10 miles	
	Cases n=25	Controls n=292	Cases n=25	Controls n=292	OR (95 % CI)	Adjusted OR[1] (95 % CI)	OR (95 % CI)	Adjusted OR[1] (95 % CI)
EDM	7 (22.5)	48 (16)	24 (77)	244 (83)	0.97 (0.93, 1.0)	0.97 (0.93, 1.0)	1.48 (0.6, 3.6)	1.3 (0.5, 3.3)
McAdoo	9 (29)	57 (19.)	22 (70)	235 (80)	0.97 (0.93, 1.0)	0.96 (0.93, 1.0)	1.68 (0.7, 3.8)	1.5 (0.6, 3.8)
Northeastern Power Company	8 (25)	46 (15)	23 (74)	246 (84)	0.97 (0.93, 1.1)	0.97 (0.93, 1.0)	1.86 (0.78, 4.4)	1.6 (0.64, 4.0)
Tonolli Corporation	8 (25)	52 (17)	23 (74)	240 (82)	0.97 (0.93, 1.1)	0.97 (0.93, 1.1)	1.6 (0.6, 3.7)	1.4 (0.5, 3.9 0
Westwood Generation	0 (0)	9 (3)	31 (100)	283 (97)	0.99 (0.97, 1.0)	1.0 (0.96, 1.0)	-	-
Wheelabrator Frackville Energy	7 (22)	55 (19)	24 (78)	237 (82)	0.98 (0.95, 1.0)	0.99 (0.94, 1.0)	1.25 (0.5, 3.0)	0.96 (0.25,3.7)

1. Odds ratio adjusted for age and sex and frequency in %

Table 4.3 Logistic regression gene and distance to Hazardous/Industrial Site stratified by near (<= 10 miles) and far (>10 miles)

Gene	Hazardous waste site/ waste-coal power plant	Residence < =10 miles		Residence >10 miles	
		OR (95% CI)	Adjusted OR [1] (95% CI)	OR (95% CI)	Adjusted OR [1] (95% CI)
AHR					
	EDM	0.5 (0.05, 5.0)	0.05 (0.05, 6.2)	1.0 (0.3, 2.9)	1.1 (0.3, 3.3)
	McAdoo	0.4 (0.04, 3.6)	0.5 (0.05, 6.2)	1.1 (0.4, 3.3)	1.2 (0.4, 3.7)
	Northeastern Power Company	0.3 (0.04, 3.6)	0.3 (0.03, 3.7)	1.1 (0.4, 3.2)	1.2 (0.4, 3.7)
	Tonolli Corporation	0.5 (0.06, 5.3)	0.7 (0.06, 8.5)	1.0 (0.3, 2.9)	1.2 (0.4, 3.5)
	Westwood Generation	-	-	0.8 (0.3, 2.2)	0.9 (0.3, 2.8)
	Wheelabrator Frackville Energy	0.5 (0.05, 4.7)	0.5 (0.04, 5.5)	1.0 (0.3, 2.9)	1.1 (0.3, 3.2)
CY1A2					
	EDM	7.8 (0.4, 142)	20.9 (0.6, 710)	2.5 (0.7, 8.1)	3.0 (0.8, 10.7)
	McAdoo	3.4 (0.2, 42.4)	15.0 (0.6, 345)	2.8 (0.8, 9.3)	3.4 (0.9, 12)
	Northeastern Power Company	3.1 (0.2, 39.4)	5.6 (0.6,50)	2.8, (0.8, 9.2)	2.9 (0.7, 11.8)
	Tonolli Corporation	-	-	3.2 (1.0, 9.6)	4.9 (1.4, 17.0)
	Westwood Generation	-	-	2.8, (0.97, 8.2)	3.5 (1.1, 11.2)
	Wheelabrator Frackville Energy	4.0 (0.6, 26.2)	5.6 (0.6, 50)	2.0 (0.6, 8.5)	2.99 (0.7, 12.2)
CY1E2					
	EDM	2.4 (0.2, 27.4)	2.9 (0.2, 42.3)	1.1 (0.1, 9.2)	1.2 (0.1, 10.9)
	McAdoo	1.5 (0.1, 16.1)	2.1 (0.1, 30.8)	1.3 (0.1, 11.1)	1.5 (0.1, 13.6)
	Northeastern Power Company	3.0 (0.2, 37.6)	2.3 (0.1, 32.8)	1.0 (0.1, 8.6)	1.2 (0.1, 10.4)
	Tonolli Corporation	5.3 (0.7, 38.6)	7.8 (0.7, 80.8)	-	-
	Westwood Generation	-	-	1.68 (0.3, 7.9)	2.0 (0.4, 10.2)
	Wheelabrator Frackville Energy	-	-	1.6 (0.3, 37.9)	1.6 (0.3,8.3)

Table 4.3 Logistic regression gene and distance to Hazardous/Industrial Site stratified by near (<= 10 miles) and far (>10 miles) (cont'd)

Gene	Hazardous waste site/ waste-coal power plant	Residence < =10 miles		Residence >10 miles	
		OR (95% CI)	Adjusted OR [1] (95% CI)	OR (95% CI)	Adjusted OR [1] (95% CI)
GSTA1					
	EDM	0.5 (0.06, 5.1)	0.6 (0.6, 6.2)	2.8 (1.1,7.0)	2.0 (0.7, 5.2)
	McAdoo	1.1 (0.2, 6.5)	0.9 (0.1, 6.1)	2.6 (1.0, 6.6)	1.8 (0.6, 4.9)
	Northeastern Power Company	0.9 (0.1, 5.3)	3.3 (0.5, 20.6)	2.6 (0.1, 6.7)	1.3 (0.4, 3.4)
	Tonolli Corporation	1.4 (0.2, 7.9)	1.1 (0.1, 7.2)	2.4 (0.9, 6.1)	1.7 (0.6, 3.9)
	WPS	-	-	2.0 (0.9, 4.6)	1.4 (0.7, 3.9)
	Wheelabrator Frackville Energy	5.1 (0.9, 27.9)	3.3 (0.5, 20.6)	1.7 (0.6, 4.3)	1.4 (0.5, 3.9)
GSTM1					
	EDM	2.3 (0.4, 13)	2.2 (0.3, 15)	2.1 (0.8, 5.2)	2.0 (0.8, 5.2)
	McAdoo	3.1 (0.6, 16.4)	2.9 (0.4, 19.7)	1.8 (0.7, 4.7)	1.8 (0.7, 4.7)
	Northeastern Power Company	2.7 (0.5, 15.0)	2.3 (0.4, 13.6)	1.9 (0.7, 5.0)	1.8 (0.7, 4.8)
	Tonolli Corporation	2.0 (0.3, 11.0)	2.0 (0.3, 12.9)	2.1 (0.8, 5.3)	2.0 (0.8, 5.3)
	Westwood Generation	-	-	2.1 (0.9, 4.8)	2.0 (0.8, 4.6)
	Wheelabrator Frackville Energy	2.5 (0.4, 14.0)	2.3 (0.3, 14.0)	2.0 (0.8, 5.1)	1.8 (0.7, 4.7)
GSTM3					
	EDM	0.8 (0.1, 4.3)	0.9 (0.1, 5.2)	1.4 (0.6, 3.4)	1.4 (0.5, 3.4)
	McAdoo	0.6 (0.1, 3.0)	0.7 (0.1, 3.9)	1.7 (0.6, 4.5)	1.7 (0.6, 4.6)
	Northeastern Power Company	0.7 (0.1, 3.6)	0.7 (0.1, 3.6)	1.6 (0.6, 3.9)	1.5 (0.6, 3.9)
	Tonolli Corporation	0.3 (0.6, 1.9)	0.4 (0.07, 2.3)	2.0 (0.7, 5.0)	1.9 (0.7, 5.0)
	Westwood Generation	-	-	1.2 (0.5, 2.6)	1.2 (0.5, 2.7)
	Wheelabrator Frackville Energy	2.8 (0.4, 15.7)	2.6 (0.4, 16.6)	1.0 (0.4, 2.3)	1.0 (0.4,2.5)
GSTT1					

Table 4.3 Logistic regression gene and distance to Hazardous/Industrial Site stratified by near (<= 10 miles) and far (>10 miles) (cont'd)

Gene	Hazardous waste site/ waste-coal power plant	Residence < =10 miles		Residence >10 miles	
		OR (95% CI)	Adjusted OR [1] (95% CI)	OR (95% CI)	Adjusted OR [1] (95% CI)
	EDM	1.0 (0.1, 6.2)	0.7 (0.1,4.9)	0.9 (0.2, 3.2)	0.9 (0.2, 3.6)
	McAdoo	0.8 (0.1, 4.2)	0.6 (0.9, 3.8)	1.0 (0.2, 3.7)	1.0 (0.2, 4.0)
	Northeastern Power Company	1.2 (0.2, 6.8)	0.8 (0.1, 5.4)	0.8 (0.2, 3.0)	0.9 (0.2, 3.5)
	Tonolli Corporation	0.2 (0.03, 2.5)	0.2 (0.02, 2.6)	1.5 (.4, 4.7)	1.4 (0.4, 4.8)
	Westwood Generation	-	-	1.0 (0.3, 2.9)	0.9 (0.2, 2.7)
	Wheelabrator Frackville Energy	1.1 (0.1, 10.7)	0.7 (0.05,9.3)	1.0 (0.3, 3.1)	0.9 (0.2, 3.0)
NAT2					
	EDM	-	-	2.2 (0.8, 5.8)	2.0 (0.7, 5.4)
	McAdoo	-	-	2.0 (0.7, 5.4)	1.9 (0.7, 5.3)
	Northeastern Power Company	5.1 (0.5, 45.1)	4.6 (0.4, 51.9)	2.7 (1.0, 7.7)	2.5 (0.9, 7.8)
	Tonolli Corporation	-	-	2.0 (0.7, 5.4)	1.9 (0.7, 5.2)
	Westwood Generation	-	-	3.2 (1.3, 8.2)	3.0 (1.1, 7.7)
	Wheelabrator Frackville Energy	3.7 (0.4, 32.9)	4.6 (0.4, 45)	3.0 (1.1, 8.4)	2.7 (0.9, 7.3)
p53					
	EDM	0.4 (0.7, 2.2)	0.3 (0.05, 2.0)	1.8 (0.7, 4.2)	1.8 (0.7, 4.4)
	McAdoo	0.2 (0.05, 1.4)	0.1 (0.02, 1.2)	2.3 (0.9, 5.7)	2.2 (0.8, 5.7)
	Northeastern Power Company	0.6 (0.12, 2.8)	0.5 (0.09, 3.3)	1.6 (0.7, 3.9)	1.9 (0.6, 3.9)
	Tonolli Corporation	0.3 (0.07, 2.1)	0.2 (0.03, 1.4)	1.9 (0.8, 4.6)	1.9 (0.7, 4.7)
	Westwood Generation	-	-	1.2 (0.6, 2.7)	1.2 (0.5, 2.6)
	Wheelabrator Frackville Energy	2.9 (0.5, 16.2)	2.2 (0.3, 13.9)	1.0 (0.4, 2.4)	1.0 (0.4, 2.5)

1. Odds ratio (OR) adjusted for age and sex and 95% confidence intervals

Chapter 5: Summary of Findings

Objectives of this dissertation

The objective of this dissertation was to investigate associations between effects of environmental risk factors (including risks associated with aromatic and heterocyclic amines) and gene-environment interactions on MPNs and JAK2 V617F mutation. This was accomplished through three specific aims:

1. evaluate the associations between lifestyle and environmental risk factors for the most common MPNs and JAK2 V617F with a diagnosis of MPN
2. explore an interaction between known susceptible genotypes for a subset of cases and controls and potential mutagenic chemical exposures, including PAHs
3. explore the joint effects of susceptible genotypes identified in Aim 2 and distance of residence from facilities with known hazardous emissions risk of MPNs.

In our study population the cases were older (median age=71 vs. 61yrs) and more likely to be male (49% vs. 39%) compared to controls, but they were otherwise demographically similar. Our sample was overwhelmingly Caucasian and none of the cases and few (2%) controls were of Jewish ancestry. All but one control were born in the US. Two-thirds of the subjects were married at the time of the interview. More cases than controls were retired (63% vs. 42%). Our study had a low response rate. This was surprising given the media coverage and the area's commitment to request and receive the CDC funding. Only 27% of cases consented for the telephone interview and only 56 % of those also consented to the optional blood draw. Our controls were slightly better, at 41% (which is fair for a Radom Digit Dialling protocol to recruit controls), with 61% of those also consenting for the blood draw.

The first aim focused on estimation of exposure to PAH as a risk factor for MPNs with and without the JAK2 V617F mutation. We found no relationships between MPNs and diet, lifestyle behaviors with presumed exposure to aromatic and heterocyclic amines, and residential history. This is the first study, to our knowledge, to look at PAH exposure from diet and smoking and the risk of developing an MPN.

Our analysis of smoking consistently refuted the existence of a positive association, with very small numbers for cases who smoked. We found an aOR of 0.8 (95% CI: 0.4, 2.0) for all MPN cases who were heavy smokers (n=5) and for JAK2- cases alone we found an aOR of 0.2 (95% CI: 0.0, 1.4), with only one case.

Our analysis with smoking was consistently null. Our findings do not support an association between PAHs and an increased risk of getting an MPN, with the null result for exposure to cigarette smoke (rich in PAH) providing the most convincing evidence. . No other differences in smoking, home, eating char grilled beef, pork, or chicken or, and recreational activities, between the two groups.This finding of no association is important and should point future research toward better assessments of other ubiquitous mutagenic chemicals, including chemicals such as benzene. In Aims 2 and 3, exposure was to benzene was considered by including susceptible genotypes that modify exposures to benzene. We considered benzene in our study; however, we only included susceptible genotypes that modify exposures to benzene—we did not study the exposure itself except for self reported exposures.

In the second aim of this dissertation, we explored the potential gene-environment interactions in the etiology of MPNs and JAK2 V617F using a biological-pathway candidate-gene approach for mutagenic chemicals.

After studying the main effects of 14 environmentally sensitive genes, we found associations only with the *NAT2*, CYP1A2, *GSTA1*, and *GSTM3* variants, with an average of 3- to 5-fold increased odds of having an MPN. Our assumption is that a main gene effect without exposure is unlikely, based on the biological pathway for disease therefore these findings suggest gene-environment interaction.

We also had some interesting findings surrounding our analysis of variants that modify enzyme function thereby magnifying the effect of the xenobiotic substrate supporting our MPN hypothesis concerning specific genotypes typically associated with benzene exposure. We detected a 2-fold increase for CYP2E1 and GSTM1 null genotypes and an 8-fold increase for the tumor suppressor gene TP53. There is a growing body of literature suggesting hematopoietic stem-cell toxicity potential (Hirabayashi, 2005). All but one case harbored at least one deleterious (meaning environmentally sensitive) SNP 30/31, with an adjusted OR of 10 (95% CI: 1.3, 79). For having two of these four compared to none, we found an adjusted OR of 15 (95% CI: 1.7, 123). For anyone with three of the four (4/31 cases), we found an adjusted OR of 43 (95% CI: 4.0, 469).

In this study, we use known biological pathways and candidate-gene polymorphisms such as the GSTM3, CYP1A2, and NAT2 gene, these genes modify environmental exposures due to increased ability to metabolize or decreased ability to detoxify the chemical. Since this is the first study to explore genetic polymorphisms and MPNs, these findings can help target future studies where only ecological environmental exposure (not individual exposure) data is available, with both disease and genotype data. Future research in the pathways affected by the NAT2, GST, and CYP genes may be fruitful avenues to better assess potential exposures and their association with MPNs.

The joint effects of susceptible genotypes and distance of current residence from facilities with known hazardous emissions

In the final aim of this dissertation, we studied the joint effects of susceptible genotypes and distance of current residence from facilities with known hazardous emissions on risk of myeloproliferative neoplasms.

The overall age and sex adjusted odds ratio for distance to any one of the six hazardous waste/waste-coal plants generally did not detect an elevated risk of MPNs with distance to these sites. All three hazardous waste sites showed similar effects Eastern Diversified Metals with an aOR of 1.3, 95% CI: 0.5, 3,3; McAdoo Associates with aOR of 1.5, 95%, CI: 0.6,3.8; Tonolli Corporation with an aOR of 1.4, 95% CI: 0.5, 3.9). We observed the strongest effect estimate from people living less than or equal to 10 miles from the Northeastern Power Company, with an aOR of 1.6 (95% CI: 0.64, 4.0), while none were of a magnitude to raise concern.

We also examined nine genotypes that suggested a signal when we looked at each risk genotype by distance category, the effect was not consistent. It was greater in some of the genotypes but not for all, compared to looking at the distance-only model in Table 4.3. The results for AHR, GSTA1, GSTM3, GSTT1, and TP53, were fluctuating around the null with overlapping confidence intervals for distances less than or equal to 10 miles. For distances greater than 10 miles, we saw a trend with a positive association for most SNPs. These dramatic swings in point estimates of the effect may be due to differences in the number of cases and controls living less than or equal to 10 miles from any particular hazardous waste/waste-coal plants. Our study did not show any consistent relationship with MPNs and joint effects of proximity to known hazardous waste site and

waste-coal power plants and environmentally sensitive genotypes, and the results we did observe were very imprecise.

Public health significance

This research intends to help focus future research to refining exposure assessments and yield relevant insights to environmental exposures and risk of MPNs. This dissertation was designed to explore the environmental etiology of polycythemia vera, essential thrombocythemia, and primary myelofibrosis. Results from this dissertation add to the limited amount of epidemiological studies looking at environmental exposures as risk factors to MPNs. Hopefully, it will encourage further research for chemical exposures as well as susceptible genotypes for these diseases.

Appendix A:

Drexel University Case-Control Study Northeast Pennsylvania
Telephone Questionnaire Content for CATI phone script
February 2011

[Interviewer script in italics]

Greet: Hello, my name is XXX and I am calling from the Center for Health Research at Geisinger.

Hello: Map I please speak to XX?

Greeting: Hi Mr./Ms. my name is[and I am calling] from The Center for Health Research at Geisinger. I am calling to conduct the phone survey for the Drexel University School of Public Health – Polycythemia Vera Cluster research study – is this a convenient time?

If no, reschedule day and time:

Participant Name:

Last: _____ Middle: _____ First: _____

Male___
Female___

Telephone: _____

Date of Interview: /__/__/ /__/__/ /__/__/__/__/
(Month) (Day) (Year)

Standard codes for Yes/No questions:
1: Yes
2: No
997: Don't know (DK)
998: Decline to answer (Ref)
999: Missing data

Appendix A:

Introduction*: Good!* *During this interview, I will ask you some questions about yourself and your family. Some questions may ask about personal information ---- I want to remind you that all of your answers will be kept strictly confidential. The information you and others provide is very important to this study.*

For all questions, answer to the best of your ability, and if you cannot recall or are unsure of an answer, please say "I cannot remember or I don't know". Ask me to repeat any question that you didn't hear or understand. You may refuse to provide an answer to any question – we'll simply continue on to the next question.

The survey will take between 45 minutes to 1 hour to complete. Please let me know at any time if you need to take a short break.

[Note to interviewer] All questions are for the time period: after respondent left school or completed education up to the present day with the following exceptions: medical conditions history (as specified in section) and residence history (after age 21).

First, I would like to capture your e-mail address in case we need to send you information later in the study. Thank you. Now, I would like to ask you some question about your place of birth and ethnic background

1. Background and Demographics

First, I'm going to ask some questions about your place of birth and ethnic background.

1-1 What is your age?

1-2 And your date of birth? (mo/day/yr):

1-3 Where were you born?

City/State (USA) _____

OR

Country of Birth:_____ Year moved to U.S.A._____

Appendix A:

1-4. And what do you consider to be your race or ethnic group? If you belong to more than one group, please tell me all the groups you belong to.

White or European American	1
Black, African American, or African ancestry	2
Latino/Latina or Hispanic (not including European Spanish or Portuguese)	3
Native American, Alaskan native, or indigenous people	4
Asian or Pacific Islander	5
Other (specify: _____)	6
Ref	99

For 1-4: If answered "1" ask 1-5 through 1-7:
If did not answer "1", skip to 1-8

1-5. What is your father's ethnic background? [interviewer: asking for heritage/nationality, limit two (paternal grandparents)]

Verbatim:

DK	98
Ref	99

1-6. What is your mother's ethnic background? [interviewer: asking for heritage/nationality, limit two (maternal grandparents)]

Verbatim:

DK	98
Ref	99

Continue probe for:

1-7. Do you have a Jewish ancestry?

Y/N/DK/Ref

1-8. What is the highest level of education you completed?

Less than high school,	1
High School/GED,	2
Some college,	3
Bachelors degree,	4
More than Bachelors.	5
Ref	99

1-9. What is your current marital status?

Married	1
Widowed	2
Currently Single	3
Ref	99

Appendix A:

1-10. Please describe your annual household income from all sources..is it:

Less than $20,000	1
Between $20,000 and $35,000	2
Between $35,000 and $50,000	3
Between $50,000 and $75,000	4
$75,000 or more?	5
DK	98
Ref	99

Next, I would like to review the residence and job history forms that we mailed to you in the original study packet. Could you get these out now?

Section 2 Interviewer Assisted Review of Mail-out Questionnaire Forms.

2-1. Residence History

I'd like to ask you to list each primary residence you've lived in for 6 months or longer, starting with your current residence (#1) and working backwards to age 21. We will start with your current residence even if you have lived there for less than 6 months. Please give me the address information to the best of your recollection, even if you don't have a complete address or ZIP Code. Please estimate the time period if you cannot remember exact dates.

Key for residence description:
1) Single Family Home
2) Apartment
3) Condominium/Townhouse
4) Farm (please specify type: dairy, livestock, cash crops, etc.)
5) Mobile Home
6) Other (please specify)

Residence number	Time period Mo/Yr to Mo/Yr	Street Address	City/Town	State and ZIP Code	How would you best describe this residence? (see key above)

Appendix A:

1 (Current)	to				
2	to				
3	to				
4	to				
5	to				

[Interviewer: after completing each line, say, Thank you. Now still working backwards, was there another residence that you lived in more than six months, going back when you were 21 years old.

Appendix A:

Thank you. We willl continue now with a set of questions for each residence you just provided me.

Residence-specific questions (numbered consecutively for each residence)

Water and Heating sources
Current Address first:
[Note to Interviewer: Ask the following questions for each residence of respondent. Start with current residence (1), and then continue in a reverse chronological order according to respondent's residence list. Maintain number coding for each residence: 1, 2, 3, etc.]

Example: Res1_1, Res1_2 …. Res1_22; then continue with Res2_1, Res2_2 … Res2_22, etc.

Let me first ask about your [R1] current residence at [interviewer state street address]. Use "Since you moved here…" for current address questions.

[Interviewer: mark all that apply if applicable]
Res(#)_1. What is your main [source] of household water?

Municipal Water	1	Private well	2
River/lake/pond	3	Rainwater/cistern	4
Other	5		
Don't know	98	Refused	99

Res(#)_2. What type of water do you use for drinking? (*allow multiple answers***)**

Tap water - unfiltered	1
Tap water - filtered	2
Spring water/bottled water	3
Well water	4
Other	5
Don't Know	98
Refused	99

Res(#)_3. What type of water do you use for cooking?

Tap water - unfiltered	1
Tap water - filtered	2
Spring water/bottled water	3
Well water	4
Other	5
Don't Know	98

Appendix A:

Refused	99

Res(#)_4. What are the major sources of heat for this home? (*allow multiple answers***)**

Electricity	1
Natural Gas/Propane	2
Kerosene	3
Fuel Oil	4
Coal	5
Wood	6
Other	7
DK	98
Ref	99

Res(#)_5. What fuels do you use for cooking? (*allow multiple answers***)**

Electricity	1
Natural Gas/Propane	2
Kerosene	3
Fuel Oil	4
Coal	5
Wood	6
Other	7
DK	98
Ref	99

Res(#)_6. Do you store any of the following fuels in any room or basement or in an attached garage or carport?

Y/N/DK/Ref	
Gasoline	1
Diesel Fuel	2
Kerosene	3
Solvents, turpentine or paints	4
Other fuel oils	5
Verbatim: _____	
DK	98
Ref	99

[Probe for each “yes” to 6; if “no” to 6 skip to 8]

Res(#)_7. About how much do you generally store? Please estimate the size of the container(s) in ounces or gallons.

Number of ounces ____ OR
Number of gallons _____
DK 98

Appendix A:

Ref	99

Res(#)_8. Do you burn your trash or yard clippings?
Y/N/DK/Ref

Res(#)_9. [If yes to 8] How often did you burn your trask or yard clippings?

Weekly	1
Monthly/	2
Yearly (few times a year or less)	3
Don't know	98
Ref	99

<u>Location near potential contamination/hazardous sites</u>

Res(#)_10. Do you live within ½ mile (or 10 blocks) of any of the following types of facilities:

Y/N/DK/Ref to each

Appendix A:

Site	If yes, site name if known	Additional Probe
1 Dump/landfill		
2 Hazardous waste site		
3 Airport		
4 Farm		If yes, what type? (a dairy, b livestock, c cash crops, d non-working, e other)
5 Nursery/greenhouse		
6 Golf course		
7 Railroad track used by train		
8 Gas station		
9 High voltage electricity tower		
10 Incinerator		
11 Factory/industrial plant		
12 Quarry/mine		
13 Coal fired power plant		
14 Nuclear power plant		
15 Other		If yes, [verbatim] specify:

Contact with Soils and Garden

Res(#)_11 When you lived at ____ in ___, did you grow any fruits or vegetables in your yard or in a garden?
Y/N/DK/Ref
[If yes, continue to Res_12]
[If no, skip to question Res_15]

Res(#)_12 Have you eaten fruits and vegetables grown from your garden (yard)?
Y/N/DK/Ref
[If yes, continue to Res_13 and 14]
[If no, skip to question Res_15]

Res(#)_13 During the growing season, how often did you eat those home-grown fruits and vegetables?

Appendix A:

Daily/Weekly	1
Less than once per month	2
Don't know	98
Ref	99

Res(#)_14 How regularly did you wash the vegetables and/or fruit before you eat/cook them?

Never	0
Sometimes	1
Always	2
DK	98
Ref	99

Now I'm going to ask about yard and garden work for [re-state residence].

Res(#)_15 Do you work in soil in your yard (e.g., gardening, digging, building, repairing)?

Y/N/DK/Ref

[If "yes", continue with 16 and 17; If N/DK/Ref: skip to Res_18, Pesticide use]

Res(#)_16 Is the activity:

Gardening/Planting	1
Building/repairing	2
Other	3
Verbatim:	
DK	98
Ref	99

Res(#)_17 How often did you do these activities?

Weekly (or more)	1
Monthly	2
Yearly (few times a year or less)	3
Don't know	98
Ref	99

<u>Pesticide Use</u>

Res(#)_18. Have you or anyone else ever used chemicals in or around your house, yard, or garden to control weeds or pests, rodents or other pests?

Appendix A:

Y/N/DK/Ref

[Prompt examples, if needed: Pesticides or other chemicals used to kill insects, weeds, rodents or other pests. (e.g. Raid, Roundup)]

Res(#)_19. Who applied the product? (*allow multiple answers***)**

Self	1
Family member	2
Exterminator	3
Other	4
Don't know	98
Ref	99

Res(#)_20. What was its purpose? (mark all that apply)

To control plants/weeds (herbicide)	1
To control insects (insecticide)	2
To control rodents (rodenticide)	3
Other	4
Verbatim:	
DK	98
Ref	99

Res(#)_21. Where was the product used: inside the house, outside the house, or both?

Inside the house	1
Outside	2
Both	3

Res(#)_22. How often were these products used to control plants and weeds? (*repeat for each answer to Res#20 above***)**

Weekly	1
Monthly	2
Yearly (few times a year or less)	3
Don't know	98
Ref	99

Thank You. Now we'll continue with your next residence at [state street address] with the same set of questions. Please remember it's okay if you don't know or remember the answer for your previous residences.

[Note to interviewer: Repeat Questions R1-R22 for each previous residence using past tense in wording of questions]

Section 2-2. Work History

Appendix A:

Now I'm going to ask questions related to your job history form.

Are you currently employed for wages, self-employed, out of work for more than 1 year, out of work for less than one year, a homemaker, a student, retired, or unable to work?

Employed for wages	1
Self-employed	2
Out of work for more than 1 year	3
Out of work for less than 1 year	4
Homemaker	5
Student	6
Retired	7
Unable to work	8
Refused to answer	99

Now we'll review each job using the mail-out form as a guide, starting with your most recent work.

Now, I'd like you to answer the following questions to the best of your ability about **each job or occupation you held for at least 1 month since you left school (completed your education)**. Please include full-time, seasonal work, part-time, volunteer work and military service (if you worked there at least 1 month). Also include your current job, even if you have had this for less than 1 month.

Let's begin with your most recent (current) job and continue back. Please estimate the time period if you cannot remember exact dates.

a. Job number (Company Name)	b. Time period Mo/Yr to Mo/Yr	c. Job Title	d. Main Job Tasks	e. City and State of Workplace	f. Briefly describe the machines, tools and materials you used on the job	g. What did your company do at that site?
1	to					
2	to					
3	to					

Appendix A:

4	to					
5	to					
6	to					

Now I would like you to please describe any gaps in your work history or any extended periods of absence from work that were not captured in your job history, such as absences for education, illness, pregnancy, care giving (i.e. child or elder care), or retirement. [Interviewers to check for existence of gaps before completing the interview]

When? (Mo/Yr)	**How Long?**	**What was the reason for the gap or absence?**

For any of the jobs you listed, did you perform any of the following tasks or activities as part of your normal duties? Please check all that apply.

Task or Activity (place check-mark if present)	**In which of your jobs did you perform this task? List all *job numbers* that apply.**
Welding _	
Painting _	
Degreasing parts or equipment _	
Working with glues _	

Appendix A:

Working with solvents or inks _	
Working where pesticides were used _ If yes: Purpose of pesticide? (check all that apply) To control plants/weeds (herbicide) __ To control insects (insecticide) __ To control rodents (rodenticide) Other	
Working with or near diesel-powered equipment _	
Working with or around live animals _	
Firefighting _	
Working with X-ray or radioactive material _	

After work history is completed, continue with script below:

Interviewer Script: We're over half-way through the survey and done with all our residence and job related questions – thanks for your patience so far. Would you like to take a short break before we continue with the final sections?

Ok, we'll continue now with the final sets of questions.

3. Hair Dye Use

3-1. Have you used any hair coloring products?
Y/N/DK/Ref

[If no, skip to Q4]
[If yes, continue]

3-2. Is the hair coloring done yourself or by a beautician/hair stylist? (*allow multiple answers***)**

Self	1
Beautician/hair stylist	2
Other (specify)	3
DK	98
Ref	99

Appendix A:

3-3. Frequency: How often do you color your hair - Weekly/Monthly/Yearly?

Weekly (or more)	1
Monthly	2
Yearly (few times a year or less)	3
Don't know	98
Ref	99

4. Grilled Meat Intake and Local Fish Consumption

The next questions are about your grilled meat intake.

4-1 Have you ever eaten grilled, barbequed, or smoked meats or fish?
Y/N/DK/Ref
[If yes, continue to 4-2]
[If no, skip to 4-6]

KEY for table below:
Did you eat (type of meat)?
Y/N/DK/Ref

If yes, continue with A, B, C, as applicable to age of respondent

D. During which seasons (answer all that apply):

Winter	1
Spring	2
Summer	3
Fall	4
All Year	5

If yes:

Have you ever eaten [type of meat]	A)Between the ages of *21 and 40* how many meals in a *typical week/month/or year* did you usually eat (type of meat)	B)Between the ages of 41 and 60, how many meals in a *typical week/month/or year* do/did you usually eat (type of meat)	C)After age 60, how many meals in a *typical week/month/or year* have you usually eaten (type of meat)	D)Which seasons of the year did you usually eat (type of meat) Answer all that apply
4-2 grilled or barbequed beef, lamb or pork?				

Appendix A:

4-3 grilled or barbequed poultry or fish?				
4-4 smoked beef, lamb or pork such as bacon or ham?				
4-5 smoked poultry or fish such as smoked turkey or lox?				

Now let's talk about the local fish consumption.

4-6. Have you ever visited the Still Creek Reservoir?
Y/N/DK/Ref

[If yes, continue to 4-7; if no skip to 4-10]

4-7. Did you take soil from the Still Creek Reservoir to your garden or backyard?
Y/N/DK/Ref

4-8. Have you ever eaten fish from the Still Creek Reservoir?
Y/N/DK/Ref

[If yes to 4-8]
4-9. During which years did you eat fish from the Still Creek Reservoir?

From _____ to _____
DK 98
Ref 99

4-10. Have you ever eaten fish from other local creeks or lakes?
Y/N/DK/Ref

4-11. [If yes] which creeks or lakes? __ Capture creeks or lakes.

[if yes to 4-10]

Appendix A:

4-12. During which years did you eat fish from these local creeks or lakes?

From _____ to _____
DK 98
Ref 99

Thank you. Now I would like to ask you some questions about hobbies.

5. Hobbies

5-1. As an adult (21 or older), have you engaged in any hobbies on a regular basis?
Y/N/DK/Ref

5-2. If yes, what are they? [list each verbatim]

[If require examples prompt: gardening, painting, arts/crafts, auto repairs, construction projects, recreational/athletic activities etc.]

[If yes] I'll ask a couple questions about each hobby:
5-3. What year (approx.) did you begin this hobby?
5-4. What year did you stop?
5-5. How often did you pursue this hobby?

Hobby (verbatim)	**Start Year**	**End Year**	**How often:** **Weekly 1** **Yearly or less 3** **Monthly 2** **DK/Ref 98/99**

---

6. General health questions

Now I would like to ask you some general health questions.

6-1. How tall are you? (feet and inches):_____

6-2. How much do you weigh? (lbs.):_____

Appendix A:

Now I'm going to ask about specific medical conditions. If you had any of the following illnesses, your doctor would have told you its name. So, if you don't recognize the name, we will assume that you've never had it.

6-3. [For each item] Has a doctor ever told you that you have [list below]?
Y/N/DK/Ref

6-4. [For each yes to 6-3] At what age were you diagnosed?____

List:
High blood pressure
Heart Disease
Stroke
Diabetes
Arthritis
Liver Disease
Kidney Disease

Multiple Sclerosis
Lupus
Fibromyalgia
Scleroderma
Reynaud's Disorder
Ulcerative Colitis
Crohn's Disease
Parkinson's Disease

Allergies
Asthma
Chronic bronchitis
Emphysema

Blood disease
    If yes, what is the disease? [verbatim]
Cancer
    What type of cancer? [verbatim]

    If yes to Cancer:
    Were you treated with:

Chemotherapy	1
Radiotherapy	2
Immunotherapy	3
Surgery	4
Other	5
DK	98
Ref	99

Appendix A:

6-5. As an adult, have you ever been treated by a doctor or other primary caregiver for an infectious disease? (examples: hepatitis, malaria, meningitis, influenza/pneumonia, etc)

Y/N/DK/Ref

If yes, what was the infectious disease [verbatim]? _____
Age:

6-6. Have you been diagnosed by a doctor for any other medical conditions that we may have not asked you about?

If yes, verbatim:
Age at diagnosis:

6-7. Have you ever had a blood transfusion (s)?
Y/N/DK/Ref

[If yes, continue with 6-8 and 6-9; if no, skip to 6-10]

6-8. When was the first time you received blood? (At what age)
6-9. What medical condition did the transfusion treat? [verbatim]

6-10. Are you currently taking any medicines regularly?
Y/N/DK/Ref

6-11. [If yes] Which ones (please list names)?

7. Blood Relatives Medical History

Now i'd like to ask several questions about the medical history of your blood relatives. A blood elatives would include your Mother, Father, and their siblings, any Brothers and Sisters and their children, your Children/Grandchildren and Grandparents, who are living or deceased.

We are asking about your blood or biological relatives only – not step or adopted relatives

7-1. To your knowledge, have any of your blood relatives had a blood disease or blood cancer?
Y/N/DK/Ref

[If N/DK/Ref, continue to Section 8]
[If "yes", repeat 7-2 through 7-4 for each additional relative]

7-2.
Which blood relative (of respondent): allow multiple answers

Appendix A:

Mother	**1**
Father	**2**
Brother	**3**
Sister	**4**
Sibling of Mother	**5**
Sibling of Father	**6**
Parent of Mother	**7**
Parent of Father	**8**
Child of Brother or Sister	**9**
Son	**10**
Daughter	**11**
Grandchild	**12**
Other	**13**
Specify:	
DK	**997**
Ref	**998**

7-3. What type of blood disease or blood cancer?
Verbatim:
DK
Ref

7-4. At about what age was he/she diagnosed?
Age: _____ or DK

That completes our medical history section and we now have two short sections left to complete.

8. Tobacco Use
The next few questions ask about your use of tobacco products.

8-1 Have you ever smoked any tobacco product (at least one cigarette, one cigar or one paper per day for at least six months)?
1 Yes 2 No (go to 9-1) 99 Ref

[If yes] *Okay, I'm going to ask about specific types of tobacco use.*

8-2 Have you ever smoked cigarettes (at least one per day for at least six months)?
1 Yes 2 No (go to 8-3)

From	To	Filter (1)	Manufactured (1)	Average number per day

Appendix A:

year/age	year/age	non filter (2)	hand rolled (2) both types (3)	Working days	Non-working days, holidays
..... |__|__|	..... |__|__|	|__|	|__|	|__|__|	|__|__|
..... |__|__|	..... |__|__|	|__|	|__|	|__|__|	|__|__|
..... |__|__|	..... |__|__|	|__|	|__|	|__|__|	|__|__|
..... |__|__|	..... |__|__|	|__|	|__|	|__|__|	|__|__|
..... |__|__|	..... |__|__|	|__|	|__|	|__|__|	|__|__|

8-3 Have you ever smoked cigars (at least one per day for at least six months)?

1 Yes 2 No (go to 8-4)

From year/age	To year/age	Cigars (1) cigarillos (2)	Average number per day	
			Working days	Non-working days, holidays
..... |__|__|	..... |__|__|	|__|	|__|__|	|__|__|
..... |__|__|	..... |__|__|	|__|	|__|__|	|__|__|
..... |__|__|	..... |__|__|	|__|	|__|__|	|__|__|

8-4 Have you ever smoked a pipe (at least one per day for at least six months)?

1 Yes 2 No

From year/age	To year/age	Average number of pipefuls (bowlfuls) of tobacco per week
..... |__|__|	..... |__|__|	|__|__|
..... |__|__|	..... |__|__|	|__|__|

---

9. Alcohol Use

Appendix A:

9-1. Did you ever drink alcoholic beverages such as wine, beer, or other liquor - at least one drink per month for 6 months or more? [if prompted: A drink of alcohol is one can or bottle of beer, one glass of wine, or a jigger of liquor either alone or in a mixed drink.] Y/N/DK/Ref

If yes,

From year/age	To year/age	Liquor (1) Beer (2) Wine (3)	Average number of drinks per day on: Working days	Non-working days, holidays
..... |__|__|	..... |__|__|	|__|	|__|__|	|__|__|
..... |__|__|	..... |__|__|	|__|	|__|__|	|__|__|
..... |__|__|	..... |__|__|	|__|	|__|__|	|__|__|
..... |__|__|	..... |__|__|	|__|	|__|__|	|__|__|
..... |__|__|	..... |__|__|	|__|	|__|__|	|__|__|

10. Final Questions

That concludes our phone survey...

10-1. [Open-ended Question] Is there anything you want us to know that we did not ask about?
Y/N/DK/Ref
[verbatim if "yes"]

Thank you again for your time Mr./Ms./Mrs. [ ]. You will receive a gift card in the mail within the next month for your contribution to our study.

If YES on biologic specimen consent:

Now I see that you signed a consent form that you will provide a blood sample. Thank you for agreeing to do so. As it mentioned in the consent form if you are not a Geisinger patient you will be registered as one to order the blood draw.

You can have your blood draw completed at one of the following Geisinger Clinics – Frackville, Pottsville, Hazleton, and Geisinger Wyoming Valley. In addition, our research team will send you an information packet regarding your blood draw. Once you schedule and complete your draw you will receive your additional $25.00 gift card in the mail. Should you have any questions or

Appendix A:

concerns please contact a member of the study team at 1-866-630-0798 and press 2.

Which of the four clinics would you prefer to complete your blood draw?

If **Hazleton**, **Pottsville**, or **Frackville**:.

Before I transfer you do you have any questions or concerns?

Ok…Thank you for your time today. Now please stay on the line while I connect you with schedule services.

Transfer Call to Scheduling Services.

If **Geisinger Wyoming Valley**:

A member of our study team will be in contact with you to arrange for your blood draw at Geisinger Wyoming Valley. Please wait for the phone call and blood draw packet in the mail (which includes a set of labels) *prior* to completing the blood draw.

Do you have any questions or concerns?

Thank you for your time today. Have a great day/night.

11. For Interviewer only: rate the reliability you assign to this respondent

11-1. Respondent Cooperation:

Good	1
Fair	2
Poor	3

11-2. Overall, how reliable was the respondent?

Highly Reliable	1
Reliable	2
Questionable	3
Low Reliability	4

11-3. Other comments specific to this respondent: _____

Interview end time:

END

Research Proposal: Investigation of a Polycythemia Vera Cluster in Northeast Pennsylvania

INTRODUCTION

In 2008, the initial results of joint efforts by the Pennsylvania Department of Health (PADOH) and the Agency for Toxic Substances and Registry (ATSDR) of the Centers for Disease Control and Prevention (CDC) confirmed a cancer cluster in a tri-county region of Northeast Pennsylvania. These investigations found the incidence of polycythemia vera – a rare blood disease – to be 4.3 times higher within the cluster area than the comparative tri-county region ($p < 0.001$) (1). The close proximity of this primary cluster zone – located at the intersection of Carbon, Luzerne, and Schuylkill Counties – to known hazardous waste sites and materials might have played a role in the disease cluster origin. Follow-up studies in the area were proposed by a scientific advisory panel – the Department of Environmental and Occupational Health of the Drexel University School of Public Health was tasked with undertaking an epidemiological study. The proposed case-control study will be the first seeking to identify risk factors of polycythemia vera in this geographic area. The study design will be a population-based, case-control study to evaluate possible lifestyle and environmental risk factors for the family of BCR/ABL-negative myeloproliferative neoplasms (MPN): polycythemia vera (PV), essential thrombocythemia (ET) and primary myelofibrosis (PMF). These diseases differ phenotypically but commonly share the same JAK2 V617F point mutation (abbr. JAK2) that is thought partly responsible for disease initiation and/or progression (2). The etiologies of these MPN and the JAK2 mutation are currently unclear.

Corresponding subject information – upon informed consent – will be obtained through telephone questionnaires by trained interviewers at the Geisinger Center for Health Research Survey Unit (Danville, PA). The questionnaire will elicit detailed responses from MPN patients and non-MPN referents regarding various residential, employment, dietary, and chemical exposure histories. The range of cases of PV, ET, and PMF to be ascertained from the tri-county area (based on preliminary Pennsylvania Cancer Registry data 2001-2007) is 100-125. Study subjects will be recruited from the tri-county population. Approximately four additional subjects

(controls) per case will be selected to maximize power. New cases will be identified through a Pennsylvania Department of Health request to the state's Cancer Registry, and through collaborations with ATSDR. Retrospective associations with an increased risk for PV, ET, and PMF diseases will be analyzed among the cases independently of potential risk factors.

A. Researchers

Principal Investigator: Arthur L. Frank, MD PhD, Professor, Chair
Co-Investigator: Carol Ann Gross-Davis, MS, Assistant Professor

Drexel University School of Public Health
Department of Environmental and Occupational Health

Principal Investigator: Porat Erlich, PhD, Research Investigator
Geisinger Center for Health Research

B. Sponsor

The Centers for Disease Control and Prevention (CDC), Department of Health and Human Services.
CDC Research Grant Number: 1R01EH000640-01
See award details, attached.

C. Studies Involving Multiple Diseases

The proposed study investigates possible risk factors for polycythemia vera (PV), and additionally, essential thrombocythemia (ET) and primary myelofibrosis (PMF). Polycythemia vera is the primary disease of interest, and the basis of the disease cluster investigation by the CDC (1). The latter two diseases – ET and PMF – share many clinical features with PV; all three diseases fall under the 2008 WHO

classification scheme of myeloproliferative neoplasms (MPN) (18). Multiple research studies review the diseases as a group, commonly termed the Philadelphia chromosome negative (Ph-) or BCR/ABL-negative classical MPN referring to their shared genetic characteristics (4, 11, 12).

Blood samples from subjects will be tested for the JAK2 V617F mutation, which is present in over 95% of primary PV cases and 50-60% of ET and PMF cases (18). Allowing the latter two disease categories into our case definition would provide a more accurate representation of the JAK2 genetic mutation burden in the affected population (17).

The diseases are similar enough that additional protocols and/or informed consents are unnecessary – the language contained within this proposal (see research design) and the consents are applicable to all disorders under investigation, given the considerable overlap. The ability to explore associations with this group of diseases that share the JAK2 mutation may provide us the opportunity to test our hypothesis since the exact role of the JAK2 mutation in the causal model is still developing.

D. Duration

The project duration, per the CDC grant, lasts two years from the study award date of September 1, 2009.

E. Risk

The study activities present minimal risk to participants. Subjects may be uncomfortable sharing personal information over the phone for the study questionnaire. Subjects who agree to a one-time routine blood draw procedure will provide their own transportation to and from the blood clinic site. Any risks for injury related to the blood draw are rare. All data will be securely stored in a password

protected computer network and locked file cabinets. Data sharing will occur between the collaborating research institutions only.

F. Subject Recruitment and Selection

The epidemiological study is a case-control design that will involve subjects with specified diseases (cases) and controls without specified diseases.

Cases will be recruited through the Pennsylvania Cancer Registry (PCR) of the Pennsylvania Department of Health, and through an existing CDC database (see CDC letter for data sharing). Followback activities with cancer patients via the PCR no longer require physician consent – only a written first contact by mail to obtain informed consent. An application for access to protected data of the PCR will be submitted upon final IRB approval. Preliminary data predict between 100-125 total eligible cases in the population.

Case Inclusion Criteria:

1) Diagnosis of polycythemia vera, essential thrombocythemia, or primary myelofibrosis between January 1, 2001 and December 31, 2010* *and*
    a. JAK2 positive, non-MPN from ATSDR 2009 community screening also eligible
2) Born between January 1, 1921 and December 31, 1968 *and*
3) Continuous residence within tri-county region (Carbon, Luzerne, or Schuylkill County) during 2000-2008

[*July, 2011 protocol amendments: The case inclusion criteria have expanded to include the 2009 and partial 2010 PCR case-level data for the tri-county area. Originally, these data were not expected to be validated or available for Drexel's use. This expansion is also in response to low participation rates in the study.

Additionally, a request has been submitted to ATSDR for the contact information of fourteen JAK2-positive individuals with no presenting MPN that participated in the 2009 community screening. See also p. 19 case definition.]

Control Inclusion Criteria:

1) Born between January 1, 1921 and December 31, 1968 *and*
2) Continuous residence within tri-county region (Carbon, Luzerne, or Schuylkill County) during 2000-2008
3) No diagnosis of PV-related outcome

Note: The study subjects' birth year range, 1921 through 1968, screens for an approximate 40 to 80 year old age bracket during the years of case diagnosis, 2001-2010 (median age of PV is 62). Efforts to maintain consistent age distributions among cases and controls include monitoring of the control selection process, and oversampling of certain demographics, if necessary.

Control Selection:

A random sample of subjects (400-500 persons) will be selected through Random-Digit Dialing (RDD), forming the control group of the study. The RDD system obtains published residence lists for the specified counties, and residences will be contacted at random by phone to seek eligible persons interested in study participation. Contacts will continue until the tri-county population is sufficiently sampled for the desired control group size (estimated at 400-500, dependent on final case numbers). The Geisinger Center for Health Research Survey Unit will conduct the RDD recruitment phase. The exact number of controls (selected during RDD and consenting to questionnaire) will be determined, in part, by the final number of cases recruited for the study.

The recruitment phase is expected to last up to three months, beginning immediately after IRB approval of the study protocol. All subjects will receive an informed consent document detailing their requirements for participation. Subjects must return

the signed consent to the Survey Unit to participate. The phone questionnaire is required for participation; the blood draw phase is voluntary.

All subjects will receive a gift card valued at $25 after completion of the phone questionnaire; a second $25 gift card will be given to those subjects who also provide a blood sample. Incentives will be mailed by the Geisinger Survey Unit after completion of each phase.

Basic demographic information will be collected to track refusals and non-respondents to characterize potential selection bias. Non-respondents will be characterized during the RDD and consent procedures. Any differences between those subjects who consent to a blood sample and those who decline will be noted.

G. Locations

The following research locations will facilitate the study:

1. *Drexel University School of Public Health*
   1505 Race Street, Philadelphia, PA, 19102

   Tasks: Study coordination, project management, data analysis

2. *Drexel Institute for Biotechnology and Virology Research*
   Pennsylvania Biotechnology Center
   3805 Old Easton Road, Doylestown, PA 18902

   Tasks: Storage of processed blood samples

3. *Geisinger Health System*
   100 N. Academy Ave., Danville, PA 17822

a. *Geisinger Center for Health Research, Survey Unit*

Tasks:

-Selection of controls via RDD

-Administration of mail-out packages and telephone questionnaires; obtain informed consents

-Data entry and final data transfer to Drexel University upon completion of questionnaires

Primary data collection will occur at the Geisinger Center for Health Research Survey Unit in Danville, PA. The Survey Unit will administer informed consents; select a control group; package mail-out forms; conduct telephone questionnaires; mail incentives; and transfer the final electronic database to Drexel University School of Public Health.

b. *Geisinger Medical Laboratories*

Clinic sites: various CAP-accredited (College of American Pathologists) Geisinger locations throughout tri-county area

Tasks: Blood draw collection and shipping sites

4. *Mt. Sinai School of Medicine, Molecular Pathology Laboratory*

One Gustave L. Levy Place, New York, NY 10029

Tasks: Genetic testing of blood samples for JAK2 mutation

5. *Columbia University Mailman School of Public Health*

630 West 168th St, New York, NY 10032

Tasks: Process blood samples and test for genetic susceptibility markers in extracted DNA; send subset of processed samples to Drexel Institute for storage

(See appended flow chart for an overview of study collaborators.)

Note on subjects and study locations:
The questionnaire data collection component does not require any travel from study participants.
Study subjects who agree to the one-time blood draw are responsible for providing their own transportation to the clinic location(s). No repeat or recurring blood draws will be performed after the initial visit without further informed consent and IRB approval.

H. Background

Community residents of Tamaqua in Northeast Pennsylvania raised concerns to state health department representatives when a group of four unrelated cases of polycythemia vera (PV) were documented on one stretch of road. Polycythemia vera is a clonal stem cell disorder marked by the proliferation of red blood cell production and other blood abnormalities (5). The Pennsylvania Department of Health (PADOH) conducted initial investigations of cancer incidence in 2004 and found an increase in the rate of PV for the tri-county region of Carbon, Luzerne, and Schuylkill Counties (the town of Tamaqua lies at the intersection of the three counties) (1). The initial findings by PADOH were corroborated by the Agency for Toxic Substances and Disease Registry in 2006 (1).

PV is a myeloproliferative neoplasm (MPN) of the bone marrow – formerly termed myeloproliferative disorder (MPD) – characterized by an overproduction of erythrocytes and often platelets and other blood cells (5). Other MPNs include essential thrombocythemia (ET), primary myelofibrosis (PMF), and chronic myeloid leukemia (CML) (18). PV has no known cause and normally occurs in about 1 of

every 100,000 people each year in the U.S. (1). In 2005, an acquired point mutation in the Janus-activated kinase-2 (JAK2) gene was discovered, which occurred in nearly all PV patients (95%+) and about 50% of those with ET or PMF, but not in other cancers (18). The JAK2 V617F mutation activates a tyrosine kinase complex in bone marrow normally responsible for regulating blood cell production through molecular signaling; the mutation disrupts the normal inhibition of growth, thus increasing blood cell production (18). No germ-line associations with the JAK2 V617F mutation have been made, though familial clustering has been documented – a large population-based study in Sweden showed a 5-7 fold increase in risk of developing an MPN as a first-degree relative of an MPN patient (30). The median age of diagnosis for PV is 62 years, with a slight male predominance. ET and PMF are similar for age and gender frequency (7, 8). The JAK2 mutation was included as a major diagnostic criterion for PV, ET and PMF by the World Health Organization (WHO) in 2008, and was used by ATSDR to validate cases in their investigation (1, 12).

Previous ATSDR investigation of polycythemia vera in cluster area

The main area of concern as identified by the ATSDR in the 2007 study is the area around Tamaqua, which also includes the towns of McAdoo, Hometown, Still Creek, at the intersection of Luzerne, Carbon and Schuylkill Counties (2007 tri-county census population estimate of 521,002). There are a number of potential hazardous sources in this compact area including acid mine tailings and drainage, waste-coal (cogeneration plants) power plants, and nine U.S. Environmental Protection Agency National Priorities Listing (Superfund) sites (1; Appendix B for listing of tri-county Superfund sites).

The goals of the ATSDR investigation which provides the preliminary data for this study were to: 1) locate all cases of PV in Carbon, Luzerne, and Schuylkill Counties, 2) confirm the diagnosis of PV cases using medical records and the JAK2 mutation, and 3) describe the characteristics of these individuals. The ATSDR study identified a statistically significant ($p<0.001$) cluster of 15 PV cases (versus ~5 expected) within the tri-county region (see Appendix A for graphic) (1). The PV cases did not have any

jobs, leisure activities, or other factors in common that may have contributed to their disease, although the study was descriptive and thus not designed to compare cases to healthy controls nor determine the cluster cause. However, local residents believe that it is related to the numerous environmental hazards in the area, which include multiple U.S. EPA Superfund sites and waste-coal power plants. The investigation also found that a significant number of confirmed PV cases had not been reported to the Pennsylvania Cancer Registry (PCR), and that many PV cases reported to the PCR during this period did not satisfy the revised 2008 WHO diagnostic criteria for PV (1).

JAK2 V617F Mutation

The 2005 discovery of the acquired point mutation JAK2 V617F yielded insight into the genetic basis of BCR/ABL-negative MPNs (1). This recurrent mutation in JAK2 occurs in > 90% of patients with polycythemia vera and 50-60% of ET and PMF patients (9, 10, 11). As a result of the widespread availability of a quantitative molecular test for the JAK2 V617F mutation, the World Health Organization formally adopted revised diagnostic criteria for PV, ET, and PMF, which included the presence of this point mutation for diagnosis (12, 13).

To date no documented external causes of PV, ET, PMF, or the JAK2 V617F mutation, have been identified. Weak etiologic associations with radiation and benzene occupational exposures were made in smaller studies which were conducted prior to JAK2 V617F discovery (1).

Genetic Susceptibility to Toxic Effects of Exposure to PAH: A Brief Overview

Polycyclic aromatic hydrocarbons (PAHs) represent a large chemical class of ubiquitous pollutant by-products of partial combustion processes – many of which are known or suspected human carcinogens (20). The most established mechanism of PAH metabolic activation involves genes, and their variants (polymorphisms), including *CYP1A1*. *CYP1A1* has been studied for associations with many cancers – notably lung cancer – because *CYP1A1*'s role in metabolism allows for the body to

break down benzo[a]pyrene from its carcinogenic forms (25, 29). Variants of the *CYP1A1* gene disrupt the metabolic breakdown of PAHs. The susceptibility to PAH carcinogenic effects on human cells are heightened in the presence of *CYP1A1* variants, affecting aryl hydrocarbon hydroxylase, the enzymatic activity of the *CYP1A1* gene (29). In the general population the prevalence of a *CYP1A1* variant can be up to 45% (but is not consistent among different ethnicities), and these individuals are at an increased risk to PAH related illness (26). Our study base is predominantly Caucasian; based on the results of a pooled analysis of 14 case-control studies, the prevalence range of the *CYP1A1 MspI* variant in Caucasian non-smokers is between 15% and 33% (29). Using the reasoning of Mendelian randomization discussed by Smith (2004), the probability of a particular genotype should not depend on the measurement of exposure (27). This will allow us to test our hypothesis using the variation on *CYP1A1* genes to see an association of exposure to PAHs on PV, ET and PMF outcomes.

PAH Exposure

Polycyclic aromatic hydrocarbons (PAHs) are ubiquitous by-products of organic matter combustion processes. PAHs – which are represented by over 100 distinct compounds that may mix together – include chemicals that are known carcinogens in humans (20).

Primary environmental sources of PAHs include (20):

- Coal, oil and gas burning
- Cigarette smoking
- Dietary sources: intake of charbroiled or barbecued meats; grains and vegetables that have been contaminated by ambient PAHs.

While there is no specific mechanism of action for environmental carcinogens to cause polycythemia vera, PAHs represent a burdensome chemical exposure category for the geographic region, considering the extensive coal and mining activities. Mt. Sinai School of Medicine, another collaborator with the CDC/ATSDR research

activities in the tri-county area, is investigating two PAHs as possible genotoxic agents to bone marrow: benzo[a]pyrene and benzo[k]fluoranthene (19).

Additional testing of "markers of susceptibility"

While the *CYP1A1 MspI* variant is the primary polymorphism of interest for the present study, additional tests may genotype other genetic markers of susceptibility (performed on the blood samples sent to Columbia University). Proposed markers would include variants in xenobiotic metabolizing enzymes, with selections guided by the ongoing National Institute of Environmental Health Sciences (NIEHS) Environmental Genome Project (35). NIEHS maintains a database of "over 600 prioritized environmentally relevant genes", which will be reviewed for genetic markers that may support this study (35, 36). The majority of selected polymorphisms for genotyping will have a minor allele frequency of at least 5% in a Caucasian population.

Tri-County Population Overview

The tri-county region consists of Carbon, Luzerne, and Schuylkill Counties in Northeast Pennsylvania. See Appendix A for tri-county area graphic.

A. Tri-County Population:

County	**Population**	**Percentage of population: white/Caucasian**	**Median age**
Carbon	62,937	96.1%	41.4
Luzerne	311,752	94.4%	42.1
Schuylkill	147,107	95.3%	42.0
Total tri-county	*521,796*	*94.9%*	*42.0*

(Data retrieved from U.S. Census American Community Survey, 2008)

B. County Population Estimates, 2000-2007:

Carbon County:
Total Population:
July 1, 2000: 58,832
July 1, 2001: 59,207
July 1, 2004: 60,653
July 1, 2007: 63,154

Luzerne County:
Total Population:
July 1, 2000: 318,555
July 1, 2001: 315,487
July 1, 2004: 311,553
July 1, 2007: 311,982

Schuylkill County:
Total Population:
July 1, 2000: 150,087
July 1, 2001: 149,114
July 1, 2004: 146,428
July 1, 2007: 147,115

Source: U. S Census 2000; ACS 2005-2007

C. Age Brackets per Control Subject Inclusion Criteria:

For 1921-1968 year of birth range, study subjects:
In 2000: aged 32-79
In 2008: aged 40-87
In 2010: aged 42-89 (at time of study recruitment)

County	**35-79 Age Group**	**Total County Population**
2000 Census	**% of total county pop.**	
Carbon	53.9%	58,802
Luzerne	52.8%	319,250
Schuylkill	53.3%	150,336
Total	*53.1%*	*528,388*
	40-84 Age Group	

2005-2007 ACS (estimates)	% of total county pop. (estimate)	Total County Population Estimate
Carbon	51.0%	62,326
Luzerne	50.4%	311,838
Schuylkill	50.0%	146,838
Total	*50.4%*	*521,002*

Source: U.S. Census Bureau, 2000 Census; 2005-2007 American Community Survey
Note: Census data age group numbers do not exactly match 1921-1968 DOB study cohort due to differences in age intervals.

D. Estimated PV annual incidence rate (per 100,000 persons) in:

U.S. (2001-2004):	1.0
Pennsylvania (2001-2003):	1.5
Tri-County (2001-2005):	1.25 (33 cases)
Tamaqua cluster (2001-2005):	3.47 (15 cases)

(From Seaman et al. 2009)

I. Research Design

Study Overview: The Drexel University unmatched case-control study in Carbon, Luzerne, and Schuylkill Counties (tri-county area) in Northeastern Pennsylvania is designed to evaluate potential environmental risk factors associated with the cluster of PV cases, and additionally ET and PMF cases in northeast Pennsylvania. To determine the presence of effect modification, a 4-to-1 unmatched case-control study will be recruited from the tri –county area in Northeast Pennsylvania. The primary exposure assessments will be drawn from residential histories, occupational histories and lifestyle factors provided in a questionnaire phase. An optional blood draw phase for cases and controls will evaluate JAK2 mutation frequencies and select susceptibilities to environmental toxins. Specifically, the primary susceptibility gene of interest will be the *CYP1A1 MspI* variant, which disrupts the metabolic breakdown of PAHs. Variations in expression of the *CYP1A1* gene influence how PAH exposures are "activated" to disease-causing biological intermediates (29).

Source Population: The source population for the study is the tri-county population of Carbon, Luzerne, and Schuylkill Counties (tri-county area) in Northeastern Pennsylvania. In 2007, an estimated 521,002 people resided in the tri-county region. Eligible subjects include those persons born between 1921 and 1968 (inclusive) who have resided continuously in the tri-county area between 2000 and 2008. This restricted age group in the three counties is estimated to be 244,870 people.

A suitable control group from the tri-county population will be selected through Random-Digit Dialing (RDD) of retrieved residential phone records. The Geisinger Survey Unit will oversee the selection of controls. Those eligible for inclusion as controls will be asked for their mailing address to receive and review the informed consent form for participation in the study.

Data collection methods: One-time phone interview; and a one-time, routine collection of a 25-30 ml blood sample.

Power and Sample Size

We are planning a study with approximately 130 case patients and 520 control patients.

Minimal detectable effects for gene-environment interaction were assessed at 80% power. The minimal detectable effects were calculated using the program Quanto (Version 1.2.4, 2009) for the standard multiplicative interaction for a plausible range of the risky allele frequencies and a range of frequencies using data from 14 studies on genetic polymorphisms and risk of lung cancer (29, 34). We had assumed that main environmental effect and the main genetic effect were both 1.0. Therefore, at the high end of the allele frequency range examined, 27%, which is for the *CYP1A1 MspI* allele, in combination with the exposure estimate of 30% for exposure to PAH from industrial sources, the minimum detectable interaction odds ratio would be close to 1.8. Alternatively, at the same exposure estimate for PAH from industrial sources, 30%, in conjunction with an allele frequency of 15% for the *CYP1A1 MspI* allele of the gene on the minimal detectable interaction, the odds ratio is less than 2.1.

SPECIFIC AIMS

Aim 1: Estimate the effect of polycyclic aromatic hydrocarbon (PAH) exposure on polycythemia vera, essential thrombocythemia, and primary myelofibrosis risk in the tri-county area, Northeast Pennsylvania.

Aim 1a: Estimate the effect of exposure to polycyclic aromatic hydrocarbons on polycythemia vera, essential thrombocythemia, and primary myelofibrosis risk from occupational and residential sources, diet, and lifestyle behaviors.

Aim 1b: Estimate the effect of exposure to polycyclic aromatic hydrocarbons measured by distance from Co-generation Power Plants on polycythemia vera, essential thrombocythemia, and primary myelofibrosis.

Aim 2: Explore whether the effects of exposure to polycyclic aromatic hydrocarbons are modified by polymorphisms in the *CYP1A1* gene.

Aim 2a: Investigate the effects of exposure to PAH using *CYP1A1 MspI* gene as a proxy for exposure on polycythemia vera, essential thrombocythemia, and primary myelofibrosis risk.

Aim 2b: The association will be stronger for people with the enzymatic activity of *CYP1A1 MspI* single nucleotide polymorphism compared to people who do not have the environmental susceptibility gene.

Aim 3: Explore the role of genotypes susceptible to chemicals on the risk for PV, ET and PMF, in the Tri-County area.

Aim 3a: The association will be stronger for people with the susceptible genotypes compared to people who do not have the genotype.

METHODS

Study Design We propose to construct an unmatched case-control study. All cases must have received a diagnosis of PV, ET, or PMF between January 1, 2001 and December 31, 2010 (exception is group of JAK2 positive, MPN negative individuals from 2009 ATSDR community screening as separate case series – request submitted in July 2011). Four controls per case will be recruited via Random Digit Dialing (RDD) within Carbon, Luzerne, and Schuylkill Counties, meeting inclusion criteria for age and residence time without diagnosis of an MPN at time of interview. The primary outcomes will be whether or not an individual was diagnosed with PV, ET or PMF between 2001 and 2010 in the tri-county area.

A suitable control group from the tri-county population will be selected through RDD of retrieved residential phone records. The Geisinger Survey Unit will oversee the selection of controls. Those eligible for inclusion as controls will be asked for their mailing address to receive and review the informed consent form for participation in the study. The population in the tri-county area is predominantly Caucasian at >95% and roughly equal for males and females. The eligibility criteria for inclusion as controls are:

1) Born between 1921-1968 (inclusively) *and*
2) Continuous residence within tri-county region (Carbon, Luzerne, Schuylkill Counties) during 2000-2008

Phone questionnaires will be administered to both cases and control to obtain details on residence and job histories, water sources, chemical exposures, and other lifestyle factors, including a brief medical and family medical history. The questionnaire phase will be followed by an optional blood draw, to test for presence of the JAK2 mutation and for toxicologically relevant genetic polymorphisms, including the variant of *CYP1A1 MspI* gene.

Data for the individual exposure measurements are not available – therefore ecological analyses of ambient environmental exposures will be applied and modeled appropriately based on the data available through the efforts of an ATSDR project to assemble all the available information into a data warehouse for the PV partners. Included in the environmental sampling are air, water, and soil testing parameters that provide emissions and ambient (or modeled) data by time and geography. Of note, ambient data from the EPA's National Air Toxics Assessment (NATA) Results and National Emission Inventory (NEI) and drinking water data from EPA Safe Drinking Water Information System (SDWIS) are available in the Polycythemia Vera (PV) Data Warehouse relevant to the tri-county area (ATSDR, pers. comm.).

Questionnaire Methods: A questionnaire will be administered by phone to all participants to collect information on lifestyle behaviors as well as detailed residential history, chemical exposure history and focused dietary history of homegrown and charred food. The phone interview will be preceded by mail-out forms for subjects to complete their residence and job histories to be used as reference guides during the phone interview.

The required questionnaire phase will include:

1) Structured and validated telephone questionnaire (~45 min. – 1 hour) to be administered by the Geisinger Survey Unit to all study participants
    a. Mail-out questionnaires on residence and employment histories sent with initial study packet for interview preparation

Upon consent, all eligible participants will be phone-interviewed (Computer Assisted Telephone Interview, CATI) by the Geisinger Survey Research Unit whose staff are trained to administer the questionnaire in a standardized, unbiased manner. Interviewers will be initially blinded to case/control status of participants. The questionnaire will elicit detailed information regarding demographic characteristics, health behaviors, socioeconomic information, residential and employment history, and any past exposures to chemicals and other hazardous materials. Residential and occupational histories will be reviewed in advance of the interview by respondents as preparation for relevant interview questions. The interview is expected to last between 45 minutes and one hour; a gift certificate incentive of $25 for participation will be offered. Responses will be stored in an electronic database at Geisinger, transferred to Drexel University upon completion, and converted to SAS statistical software.

Blood draw (optional for subjects) includes:

1) One-time peripheral venipuncture (routine blood draw) of 25-30ml of blood (about 2 tablespoons)

Samples include (maximum volume listed):

1) 10ml for JAK2 testing at Mt. Sinai School of Medicine
2) 10ml for gene susceptibility testing at Columbia University
3) 10ml for storage and possible future testing for biomarkers linked to MPN by other studies ongoing in tri-county area (PV Partners studies) – storage at Drexel Institute in Doylestown, PA

Consent to provide blood samples will not be part of the study eligibility criteria; a description with a yes/no line at the end of the consent form – in addition to a second incentive ($25 gift card) for this phase – are used for blood draw recruitment. This will allow us to evaluate if those consenting to a blood draw differ significantly from non-consenters. The blood samples will be stored at the Drexel Institute for

Biotechnology and Virology Research in Doylestown, PA; Mt. Sinai and Columbia Universities will perform the genetic tests.

Blood samples: Geisinger Medical Laboratories clinics will be available as collection sites for blood samples from volunteering study participants. EDTA soft plastic tubes will be used to draw blood from each subject (25-30 ml, maximum). Collected tubes will be transported to the central Geisinger Medical Laboratories facility for shipping within 24-36 hours of collection. Tubes will be shipped with cold-storage packs to Mt. Sinai School of Medicine (New York, NY 10029) and Columbia University (New York, NY 10032).

Testing Sites

As part of the overall ATSDR research portfolio of studies occurring in the tri-county region (PV Partners studies), Drexel will have its JAK2 blood samples tested at Mt. Sinai, which serves all research groups for quantitative JAK2 mutation analysis (see Mt. Sinai validation protocol, attached). The remaining blood samples (20-25 ml) will be shipped to Columbia University for processing (DNA extraction), testing for genetic susceptibility markers, and shipping back to Drexel for storage of extra samples for future testing. A separate incentive of $25 will be offered for the blood draw.

General Approach:

The primary goal is to determine the presence (or absence) of increased odds of disease using a 4-to-1 unmatched case-control study conducted in the tri –county area in Northeast Pennsylvania (Carbon, Luzerne, and Schuylkill Counties). The primary exposures under study will be environmental exposures assessed by questionnaires eliciting residential histories, occupational histories and other lifestyle/behavioral factors including diet. The variable associations will be evaluated using various statistical methods. Analysis will control for potential confounding variables as appropriate. Available clinical and environmental data will be reviewed with the

study team prior to analysis to determine which variables may warrant exclusion or should be considered as potential confounders.

Study Subjects

Case Definition: The cases will be obtained from the tri-county area through the Pennsylvania Cancer Registry and ATSDR records. Clinical criteria for classical BCR/ABL-negative MPN: polycythemia vera (PV), essential thrombocythemia (ET), and primary myelofibrosis (PMF), diagnosed between 2001-2010 during residence in tri-county area. A proposed case series [request submitted to ATSDR for patient information in July, 2011] includes the fourteen JAK2 positive persons identified in the 2009 ATSDR community screening (16)

Primary outcome variables will be classified as 1) PV 2) ET and 3) PMF. Primary outcomes will be determined via ICD-O codes from Pennsylvania Cancer Registry (PCR) for the tri-county area data sources. All potential cases will be reviewed by expert panel members for a blinded medical review using WHO diagnostic criteria.

Tri-County MPN Reports to the PA Cancer Registry (PCR) (*Estimates provided by PCR)*

PCR Data	PV	ET	PMF	Total
2001-2007	128**	48	23	101*
2006-2007	30	N/A	N/A	N/A

** The total represents ET and PMF cases from 2001-2007 and PV cases from 2006-2007 reported to the PCR.*

***This 2001-2007 incidence figure (128) is not valid, as determined by the previous ATSDR investigation (1), and is not included in the total estimate. Reporting and diagnostics prior to 2005 (JAK2 mutation testing available) may be unreliable. Thus, the expert panel will review all unverified cases from this period. (The 2008 estimates are not yet available to the researchers.)*

N/A = Not applicable

ICD-O-3 Codes used for reporting MPN cancer cases to PCR:

Disease	ICD-O-3 code
Polycythemia vera	M-9950/3
Essential thrombocythemia	M-9962/3
Primary myelofibrosis	M-9961/3; M-9931/3

(Provided by PADOH, February 2010)

Expert Panel:

An expert panel has been formed by the Pennsylvania Department of Health to systematically review and confirm potential cases for study inclusion. Expert panel members will review de-identified medical charts obtained with the permission of patient and physician. Potential cases will be recruited through requests by the PCR. Potential cases that were previously confirmed by an expert panel will not be reviewed again. (1) Diagnostics must agree with the 2008 WHO MPN Classification scheme (PV, ET, PMF) for inclusion as cases. The four expert panel members are:

Hamid A. B. Al-Mondhiry, MD
Professor of Medicine – Hematology Division of Hematology-Oncology
The Pennsylvania State University College of Medicine
The Milton S. Hershey Medical Center

Emmanuel Besa, MD
Professor of Medicine
Hematologic Malignancies
Thomas Jefferson University
Jefferson Medical College

Samuel M. Lesko, MD, MPH
Medical Director and Director of Research
Northeast Regional Cancer Institute
University of Scranton Campus

Albert Thomas Quiery Jr., MD
Director, Hematology/Oncology Geisinger Medical Center
Geisinger Medical Center Hematology/Oncology

Summary of Variables

Data Available

Potential Covariate	**Variable Type**	**Data Source**
Age	Continuous and Categorical	Questionnaire
Gender	Binary	Questionnaire
Drinking Water Source	Categorical	Questionnaire
Burning trash on property	Binary	Questionnaire
Smoking	Categorical	Questionnaire
Eating local fish	Binary	Questionnaire
Age at diagnosis	Continuous and Categorical	Questionnaire
Proximity to Coal Power plants	Continuous and Categorical	Questionnaire
Proximity to Superfund sites	Binary	Questionnaire
Family history of MPN disease	Binary	Questionnaire
Occupation classification	Categorical	Questionnaire
PAH dietary exposure	Continuous and Categorical	Dietary Questions (grilled and barbecued meats)
JAK2 V617F	Binary	Blood sample

Summary of Exposure Variables		
Exposure Period	Parameterization	Dose Response

Drinking Water Source testing	Any vs. none	Quantiles and/or continuous
Air exposure by Cogeneration Plants	Geographical	Quantiles
Lifestyle exposures	Questionnaire	Quantiles and/or continuous
Genetic polymorphisms	Binary	Blood sample

General approach

Exploratory analysis will be conducted on all data collected from a questionnaire developed by Drexel University and administered to both cases and controls. Analysis will be conducted on data collected from the questionnaire, genotyping, and regional environmental data. Descriptive analysis will be conducted on characteristics for the study population. For categorical variables, frequency distributions will be conducted for exposure variables, genotype information, and covariates in case and control groups. For continuous exposure variables, central tendency and dispersion of the distribution will be conducted. Bivariate analysis of covariates will be used to find unadjusted odds ratios for cases and referents. Unconditional logistic regression will be used to estimate adjusted odds ratios and 95% confidence intervals.

Data Analysis

Data Analysis for Aim 1a

Analysis of case-referent data utilizes standard logistic regression; because we have multiple outcomes with one referent group, a polytomous regression approach will

also be used as this may be more efficient. A polytomous regression model will be explored as there is no inherent ordering of the outcome categories. A stratified analysis will be conducted to explore confounding and effect modification. Logistic regression will be used to estimate adjusted odds ratios and 95% confidence intervals. Categories for exposure variables will be created based on the distribution of exposure and time from historical environmental data, and data collected from the questionnaire. Potential confounders will be explored using a backward stepwise method. If the individual variable changes the effect estimate odds ratio by greater than 10%, it will be retained in the model. All variables removed will be added individually back to the model to check for joint confounding. Classic covariate adjustment of confounding risk factors will be performed, however no known strong confounders have been identified in the literature. Exposure and outcomes will be modeled using multiple logistic regression model and polytomous logistic regression. A classical error model will be used to assess measurement error in the model, and a sensitivity analysis on variables included in the final model will describe the impact of measurement error on the calculated odds ratios.

Data Analysis for Aim 1b

The overall goal is to develop a better appreciation for validity of proxy exposure measures used in air pollution exposure modeling, using community exposures to PAH from multiple coal-powered plants as a motivating example. Analysis will create a model of air pollution in a town with 3-5 coal plants nearby, each with different emission rates (Q) and volumes (V) and distances to center of town (D) (a circle). Each person in town will live some distance from the city center (r) and therefore the plant (d). Assume between-person variance, systematic increase in PAH exposure in smokers, and influence of sources on PAH all act together to produce exposure X. SAS will simulate one huge dataset of 1,000,000 records and study what happens when sampling variation does not matter, i.e. what happens "in probability", with infinite sample size.

Approach for Aim 2. Explore whether PAH exposure (overall and in specific exposures from questionnaire) are modified by polymorphisms in the *CYP1A1 MspI* polymorphism.

Data Analysis for Aim 2

Using genotype as an 'instrument', it is possible in principle to distinguish between causal and non-causal explanations of a biomarker–disease association, but classical methods for instrumental variables are rarely used (37). Classical methods of analysis of instrumental variables have some limitations; however, these methods in theory can distinguish between non-causal and causal associations with the *CYP1A1 MspI* genotype and PV, ET, and PMF. To use the Mendelian Randomization context, it is crucial to have a robust instrument and to use instrumental variable analysis to accurately make inferences from the results (38). In the analysis of gene-environment (PAH) interactions, a variable for susceptible genotype will be entered into the disease model as the instrumental variable. The relationship we are interested in testing is that genetic susceptibility is unique to PAH exposure. Therefore only if there were enough PAH exposure in this population would we see an association between genetic susceptibility and risk of PV, ET or PMF. Using this model is appropriated because the exposure in the study area is widespread and there is no evidence to support that the genetic susceptibility of the *CYP1A1* gene alone is associated with the outcomes of interest (28). To test whether factors are causally related to diseases, the classical method for instrumental variables, the two-stage least-squares (2SLS), will be used. There are two stages in the computation: in the first stage, a new variable from the instruments is created by a phenotype that is estimated for each genotypic group; in stage two, this new variable is used in place of the problematic variables X (PAH exposure) in the final regression (37).

Using this model is appropriated because the exposure in the study area is widespread and there is no evidence to support that the genetic susceptibility of the *CYP1A1* gene alone is associated with the outcome of interest (28).

Data Analysis for Aim 3

This gene-only analysis will use logistic regression to estimate adjusted odds ratios and 95% confidence intervals. Classic covariate adjustment of traditional confounding risk factors will be performed using multivariate methods. Exposure and outcomes will be modeled using multiple logistic regression model and polytomous logistic regression. For genes with more than one functional SNP, dummy variables will be created and used in Aim 2 and Aim 3. This gene-only analysis will use logistic regression to estimate adjusted odds ratios and 95% confidence intervals. Covariates will be added to this model if they were retained in the final model in Aim1. Genotypes and outcomes will be modeled using multiple logistic regression model and polytomous logistic regression. In addition to logistic regression, case-only approach will be used to test for departure from multiplicative effects on disease. Spearman and Peirce correlation coefficients will be calculated to determine the assumption of independence for gene and exposure. The assumption will be tested to affirm that this assumption is met (39). This approach is more statistically efficient than conventional approaches of multiplicative interaction but requires an assumption of independence between the exposure and genotype, which should be easily met here (39). Here logistic regression is limited to the PV, ET, and PMF cases only with the exposure alone, the genotype alone, and their joint effects. This is then compared to the odds ratio among the control subjects only. For genes with more than one functional SNP, dummy variables will be created and used in Aim 2 and Aim 3. An accurate measure of genotype would mean an expected decrease in classification error, which would increase our ability to detect an association with the disease (40). So looking at genotypes with 30-50% allele frequencies (exposure in population) with high sensitivity and specificity for the test makes it much easier to increase the probability of detecting an association with the disease (40).

Strengths

The case-control study will benefit from being the first of its kind in the tri-county area. The population based sampling approach used to select our controls in conjunction with genetic and environmental data make this analysis advantageous. Results from this analysis will add to the limited amount of literature on gene-environmental interaction in regards to environmental exposures and risk for PV. The generation of genotype data may provide an important insight on environmental exposures for PV in the context of environmental heterogeneity.

Limitations

The proposed study has several limitations. The principal limitation of the study is the maximum number of cases that are expected to be recruited from the area (approximately 125 MPN cases). Another limitation of the study design is its inability to detect small relative risks even when associated with widespread exposure. This study would be unable to detect very small relative risks <2.0 even if the exposure is widespread and large numbers of cases of cancer are occurring in the population. The study will suffer from the difficulty in small studies of detecting increased relative risks with the comparatively low power of 80%, even when maximizing the number of control per case to five. The latency period between possible exposures related to the development of an MPN is currently unknown, which complicates identification of potential risk factors when the timeframe of effective exposure is not clearly defined. Another possible limitation of this study is the potential bias from misclassification of exposure and measurement error in exposure assessment for controls. Sensitivity analysis and imputation techniques will be incorporated to quantitatively assess impact of measurement error and other exposure covariates on the overall effect estimate.

Ethical Aspects

Informed consent

Written informed consent will be obtained from all individuals participating in this study.

Involvement of Human Subjects

The research activities described in this proposal (questionnaires and optional blood draw) will be administered to consenting subjects. The study population includes men and women living in the tri-county area aged 42 to 89 years old. No exclusions are made based on gender or ethnicity; over 95% of the tri-county population is white/Caucasian. In this analysis we will be using blood samples, which will not be de-identified.

Level of Review:

The level of review for the case-control analysis will go into an expedited IRB review at Geisinger Health System, which will also rely on the Drexel University IRB.

J. Risks and Benefits to the Subject

Potential risks and protective measures

Participants may be uncomfortable answering personal questions on paper or over the telephone. Sensitive questions are not included in this protocol. All answers will remain private except to the involved researchers. The phone interview will take up to an hour to complete. No personal data will be used or shared outside the scientists working on this study, unless expressly warranted.

The specimen collection presents minimal risk to the participants. The one-time blood draw is a routine non-fasting procedure to collect 25-30 ml of blood from each participant. All blood collections will be performed at licensed phlebotomy clinics with appropriate supervisions and licensure.

The collaborating clinics will refuse collection from any participant who has undergone chemotherapy within the prior two months, is a hemophiliac, or presents

with no acceptable peripheral blood draw site (e.g. rashes, burns, wounds on arms) (23).

For the registry-based activities, personal identifiers are retained for the purposes of the data set linkages. Study results will always be presented in aggregate form, thereby further preventing identification of individual subjects. Secondly, because a biological sample may be collected, this could be of moderate concern to the participant, but study results will always be presented in aggregate form. No clinical diagnosis will be conducted on any stored sample beyond the initial screening for the JAK2 V617F mutation, which is not a diagnosis of a disease.

Participation in the study may involve unforeseen risks. In the uncommon event that excess bleeding results from the blood draw, the subject will be treated at the time of the blood draw. If any problems occur after that time, it will be the subject's responsibility, at their own cost, to seek any treatment or evaluation. Any breach of confidentiality will be referred to the Geisinger IRB and Drexel Office of Regulatory Research Compliance (ORRC).

All potential risks listed above are also detailed in the consent form.

Benefits

There are no immediate benefits to the study participants other than the cash equivalent incentive, which is modest. Controls that receive a positive genetic test for the JAK2 mutation will be notified and may be eligible to participate in a separate prospective study designed to track and detect changes in a subject's hematological profile. The local community may benefit from our improved understanding of the causes of these diseases. This information may lead to possible intervention and prevention strategies for the local community. These benefits are believed to outweigh any minimal risk to individual participants.

K. Protection of Subject Privacy and Confidentiality

1. The expert panel may review hematological medical records retrieved from the Pennsylvania Cancer Registry (PCR) through the University of Pittsburgh case ascertainment study, co-occurring in the tri-county region. The PCR documents standardize diagnostic data relevant to the diagnosis. Data from PCR records may be viewed by the Drexel investigators. No PHI directly from private medical records will be retained or recorded into investigator data collection sheets – the expert panel, under authority of the Pennsylvania Department of Health will have access to the medical record, but the investigators will not.

2. Research protocols and methodologies will be maintained on-site at Drexel University. All data will be de-identified through numeric coding; codes and data will be stored in separate files, for up to ten years following completion of the study.

With the exception of disclosure to DHHS, IRB, or collaborating research partners, data will not be linked back to individual subjects. Upon completion of a data sharing agreement, the University of Pittsburgh Graduate School of Public Health (PI: Jeanine Buchanich, PhD), in partnership with the Pennsylvania Department of Health, will have access to data sharing as needed to facilitate the aims of their study, also in the tri-county area, to prevent unnecessary participation burdens on the study population. Only phone questionnaire data for cases consented and enrolled in both the University of Pittsburgh case ascertainment study and the Drexel University case-control study may be transmitted from Drexel to Pittsburgh. Drexel University will be responsible for reviewing and matching case lists to determine which individuals (cases only) are common to the concurrent studies. No data linked to Drexel's control series may be transmitted. Transmitted data may include demographic information and relevant environmental and occupational history sections elicited by the phone survey. The data sharing is intended to limit redundant study activities (as questionnaires) that otherwise would have occurred twice for the same individual. The University of Pittsburgh case ascertainment study is funded by the Pennsylvania Department of Health.

3. No audio/video taping will be recorded.

L. Potential Conflict of Interest

The Principal Investigator, Arthur L. Frank, declares no potential conflict of interest. Co-investigator Carol Ann Gross-Davis reports her current status as a paid employee of the U.S. Environmental Protection Agency.

M. Compensation for Participation

Study participants will be mailed a gift card voucher worth $25 upon completion of the phone questionnaire interview. A separate incentive, also $25 gift card voucher, will be provided to the study participants who provide a specimen collection (blood draw). The maximum total incentive value for an individual participant is $50; incentives are to be issued independently of each other.

No additional compensation for a participant's contribution to the study will be made.

References:

1. Seaman, V. et al. Use of Molecular Testing to Identify a Cluster of Patients with Polycythemia Vera in Eastern Pennsylvania. *Cancer Epidemiology Biomarkers & Prevention*, 2009. 18(2): p. 534-540.
2. Moliterno A. Phenotypic variability within the JAK2 V617F-positive MPD: The roles of progenitor cell and neutrophil allele burdens. *Exp Hematol*., 2008 November; 36(11): 1480–1486.
3. Baxter EJ, Scott LM, Campbell PJ, East C, Fourouclas N, Swanton S, Vassiliou GS, Bench AJ, Boyd EM, Curtin N, Scott MA, Erber WN, Green AR. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet* 2005 March 19; 365(9464):1054–1061. [PubMed: 15781101]
4. Levine RL, Wadleigh M, Cools J, Ebert BL, Wernig G, Huntly BJ, Boggon TJ, Wlodarska I, Clark JJ, Moore S, Adelsperger J, Koo S, Lee JC, Gabriel S, Mercher T, D'Andrea A, Frohling S, Dohner K, Marynen P, Vandenberghe P, Mesa RA, Tefferi A, Griffin JD, Eck MJ, Sellers WR, et al. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell* 2005 April; 7(4):387–397. [PubMed: 15837627]

5. James C, Ugo V, Le Couedic JP, Staerk J, Delhommeau F, Lacout C, Garcon L, Raslova H, Berger R, naceur-Griscelli A, Villeval JL, Constantinescu SN, Casadevall N, Vainchenker W. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature* 2005 April 28; 434 (7037):1144–1148. [PubMed: 15793561]
6. Berlin N. Diagnosis and classification of the polycythemia. *Semin Hematol* 1975; 12:339–51.
7. Finazzi G, Barbui T. How I treat patients with polycythemia vera. *Blood* 2007; 109:5104–11.
8. Mesa RA, Verstovsek S, Cervantes F, et al.*;* for the International Working Group for Myelofibrosis Research and Treatment (IWG-MRT). Primary myelofibrosis (PMF), post polycythemia vera myelofibrosis (post-PV MF), post essential thrombocythemia myelofibrosis (post-ET MF), blast phase PMF (PMF-BP): consensus on terminology by the International Working Group for Myelofibrosis Research and Treatment (IWG-MRT). *Leuk Res* 2007; 31:737–40.
9. James C, Delhommeau F, Marzac C, et al. Detection of JAK2V617F as a first intention diagnostic test for erythrocytosis. *Leukemia* 2006*;* 20:350–*53.*
10. Kralovics R, Passamonti F, Buser AS, et al. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med* 2005*; 352:1779–90.*
11. Levine RL, Wadleigh M, Cools J, et al. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell* 2005*; 7:387–97.*
12. Tefferi A, Thiele J, Orazi A, et al. Proposals and rationale for revision of the World Health Organization diagnostic criteria for polycythemia vera, essential thrombocythemia, and primary myelofibrosis: recommendations from an Ad hoc International Expert Panel. *Blood* 2007*;* 110:1092–*97.*
13. Tefferi A, Pardanani A. Evaluation of "increased" hemoglobin in the JAK2 mutations era: a diagnostic algorithm based on genetic tests. *Mayo Clin Proc* 2007*;* 82:599–*604.*
14. U.S. Environmental Protection Agency. Pennsylvania Superfund Sites. Retrieved from: <http://www.epa.gov/reg3hwmd/super/pa.htm> Accessed July 2010.
15. U.S. Census Bureau, 2008 American Community Survey. Retrieved from: <http://www.census.gov/acs/www/>. Accessed May 2010.
16. Agency for Toxic Substances and Disease Registry. "Community Health Screening Report: Community Health Screening for JAK2 (V617F) Mutation. Luzerne, Schuylkill, and Carbon Counties, Pennsylvania." May 11, 2010. Retrieved from: <http://www.atsdr.cdc.gov/sites/polycythemia_vera/index.html> Accessed May 2010.
17. Malaysz J, Crisan D. Correlation of JAK2 V617F mutant allele quantitation with clinical presentation and type of chronic myeloproliferative neoplasm. *Annals of Clinical & Laboratory Science* 2009, 39:345-350.
18. Spivak J. Narrative review: thrombocytosis, polycythemia vera, and JAK2 mutations: the phenotypic mimicry of chronic myeloproliferation. *Annals of Internal Medicine*, 2010, 152:300-306.

19. Seaman et al. A multidisciplinary investigation of a polycythemia vera cancer cluster of unknown origin. *International Journal of Environmental Research and Public Health*, 2010, 7:1139-1152.
20. Martí-Cid et al. Evolution of the dietary exposure to polycyclic aromatic hydrocarbons in Catalonia, Spain. *Food and Chemical Toxicology*, 2008, 46:3163-3171.
21. Breast Cancer and the Environment on Long Island. Section C: Residential History Questionnaire. Retrieved from <http://epi.grants.cancer.gov/LIBCSP/projects/Questionnaire.html> Accessed July 2010
22. Pennsylvania Department of Health (PADOH). Pennsylvania Cancer Registry (PCR). ICD-O-3 codes and estimated incidence rates for MPN in tri-county area.
23. National Health and Nutrition Examination Survey (NHANES). Laboratory Components 2009-2010. "Environmental Health Profile". Retrieved from: <http://www.cdc.gov/nchs/nhanes/nhanes2009-2010/questexam09_10.htm> Accessed May 2010.
24. Gammon et al. Polycyclic aromatic hydrocarbon-DNA adducts and breast cancer: a pooled analysis. *Archives of Environmental Health*, 2004, 59: 640-649.
25. McManus ME, Burgess WM, Veronese ME, Huggett A, Quattrochi LC, Tukey RH. Metabolism of 2-acetylaminofluorene and benzo[a]pyrene and activation of food-derived heterocyclic amine mutagens by human cytochrome P-450. *Cancer Res* 1990 50:3367-3376.
26. Calabrese E. Pollutants and high risk groups: The biological basis of increased human susceptibility to environmental and occupational pollutants. New York, NY: John Wiley and Sons, 1978.
27. Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2004; 33(1):30-42.
28. Burstyn I, Kim H-M, Yasui Y, Cherry NM. The virtues of a deliberately misspecified disease model in demonstrating a gene-environment interaction. *Occup Environ Med* 2008; *(provisionally accepted, May 8; revised June 20, 2008)*.
29. Hung RJ et al. CYP1A1 and GSTM1 genetic polymorphisms and lung cancer risk in Caucasian non-smokers: a pooled analysis. *Carcinogenesis*, 2003 24: 875-882.
30. Landgren O et al. Increased risks of polycythemia vera, essential thrombocythemia, and myelofibrosis among 24,577 first-degree relatives of 11,039 patients with myeloproliferative neoplasms in Sweden. *Blood*, 2008, 112: 2199-2204.
31. U.S. Census Bureau. Census 2000. Retrieved from <http://www.census.gov/main/www/cen2000.html> Accessed June 2010.
32. U.S. Census Bureau. American Community Survey (ACS). 2007 ACS Data Release. Retrieved from: < http://www.census.gov/acs/www/Products/2007/index.html> Accessed June 2010.
33. Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. *Epidemiology*. 1992; 3(5):452–456.
34. Quanto. Version 1.2.4 May 2009. Downloaded from < http://hydra.usc.edu/gxe/> Accessed July 2010.
35. National Institute of Environmental Health Sciences (NIEHS). Environmental Genome

Project. Retrieved from <http://www.niehs.nih.gov/research/supported/programs/egp/>. Accessed August 2010.

36. University of Utah Genome Center. “GeneSNPs”. Retrieved from <http://www.genome.utah.edu/genesnps/> Accessed August 2010.

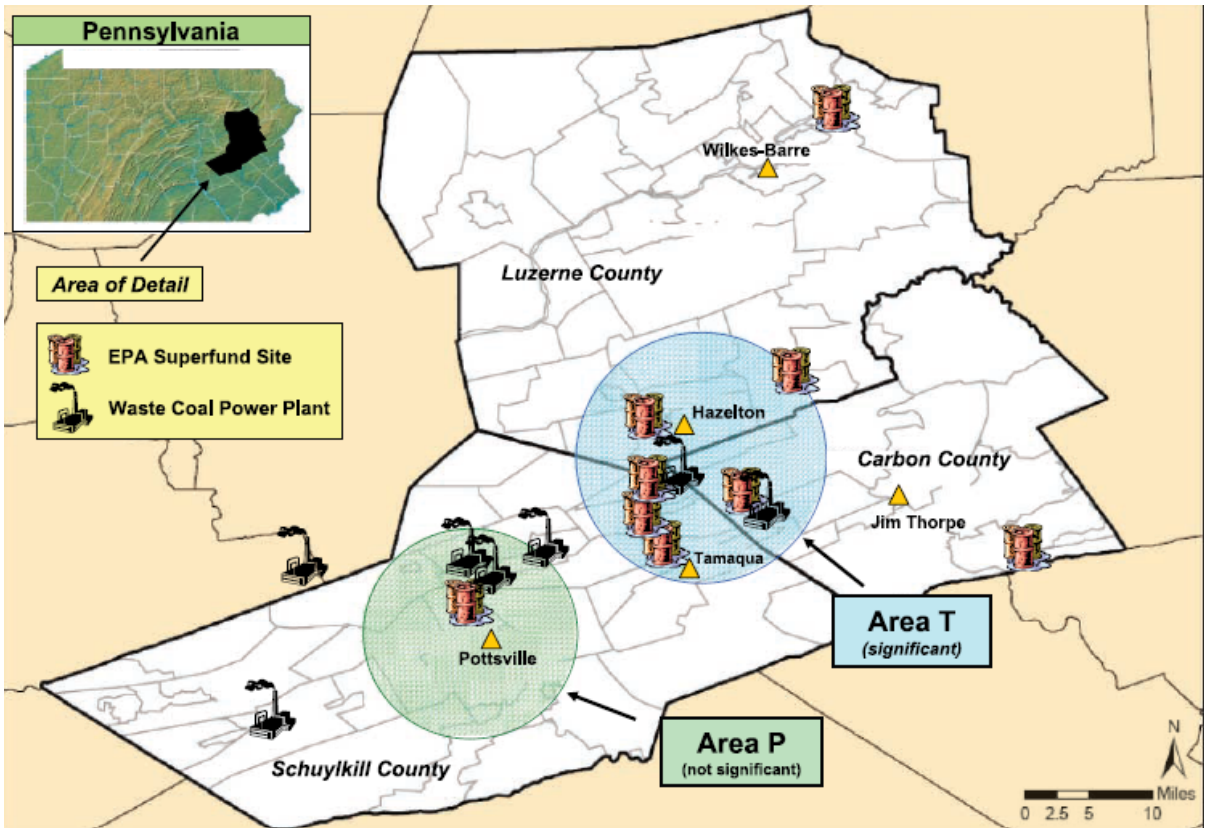
37. McKeigue PM, Campbell H, Wild S et al. Bayesian methods for instrumental variable analysis with genetic instruments ('Mendelian randomization'): example with urate transporter SLC2A9 as an instrumental variable for effect of urate levels on metabolic syndrome. *International Journal of Epidemiology* 2010, 39: 907-918.

38. Lawlor DA, Windmeijer F, Smith GD. Is Mendelian randomization ‘lost in translation?’: Comments on ‘Mendelian randomization equals instrumental variable analysis with genetic instruments’ by Wehby *et al*. *Statistics in Medicine* 2008, 27: 2750-2755.

39. Khoury MJ & Flanders WD. Nontraditional epidemiological approaches in the analysis of gene-environment interaction: case-control studies with no controls! *American Journal of Epidemiology* 1996, 144: 207-213.

40. Vineis P. A self-fulfilling prophecy: are we underestimating the role of the environment in gene-environment interaction research? *International Journal of Epidemiology* 2004, 33:945-946.

Appendix A: Graphic of Tri-County Region and Environmental Hazards. Area T represents the statistically significant cluster of polycythemia vera cases relative to the entire tri-county region.



From: Seaman et al. (2009)

Appendix B: List of Superfund Sites in Carbon, Luzerne, and Schuylkill Counties

Site Name	EPA ID	NPL Status	City	County	Zip
Palmerton Zinc Piles	PAD002395887	Final	Palmerton	Carbon	18071
Tonolli Corp	PAD073613663	Final	Nesquehoning	Carbon	18240
Butler Mine Tunnel	PAD980508451	Final	Pittston TWP	Luzerne	18640
C & D Recycling	PAD021449244	Final	Freeland	Luzerne	18224
Dallas Cleaners Site	PAN000306173	Non	Dallas	Luzerne	18612
Foster Wheeler Energy	PAD003031788	Proposed	Mountaintop	Luzerne	18707

Corporation/Church Road TCE					
Kevak Property	PAD981740129	Non	Glen Lyon	Luzerne	18617
Lehman MTBE	PA0000057471	Non	Lehman	Luzerne	18627
Moosic	PA0002008506	Non	Avoca	Luzerne	18641
Tranguch Gasoline	PA0001409671	Non	Hazleton	Luzerne	18201
Valmont TCE (Former Valmont Industrial Park)	PAD982363970	Final	West Hazleton	Luzerne	18201
Eastern Diversified Metals	PAD980830533	Final	Hometown	Schuylkill	18252
McAdoo Associates	PAD980712616	Deleted	McAdoo	Schuylkill	18237
Metropolitan Mirror & Glass	PAD982366957	Deleted	Frackville	Schuylkill	17931

Retrieved from Pennsylvania Superfund Sites:
<http://www.epa.gov/reg3hwmd/super/pa.htm> Accessed July 2010.

(NPL: National Priorities List)

Appendix C: Year Householder Moved into Unit
Source: 2006-2008 American Community Survey 3-Year Estimates, Data Profile

Carbon County

	Estimate	**Margin of Error**	**Percent**	**Margin of Error**
Occupied housing units	**25,747**	**+/-750**	**100%**	**(X)**
Moved in 2005 or later	5,517	+/-624	21.4%	+/-2.2
Moved in 2000 to 2004	5,816	+/-543	22.6%	+/-1.9
Moved in 1990 to 1999	5,991	+/-519	23.3%	+/-2.0
Moved in 1980 to 1989	2,783	+/-355	10.8%	+/-1.4
Moved in 1970 to 1979	2,361	+/-302	9.2%	+/-1.2
Moved in 1969 or earlier	3,279	+/-305	12.7%	+/-1.2

Luzerne County

	Estimate	Margin of Error	Percent	Margin of Error
Occupied housing units	**129,204**	**+/-1,450**	**99.90%**	**(X)**
Moved in 2005 or later	26,898	+/-1,155	20.8%	+/-0.8
Moved in 2000 to 2004	27,157	+/-1,309	21.0%	+/-0.9
Moved in 1990 to 1999	25,720	+/-1,097	19.9%	+/-0.9
Moved in 1980 to 1989	15,810	+/-947	12.2%	+/-0.7
Moved in 1970 to 1979	13,885	+/-723	10.7%	+/-0.6
Moved in 1969 or earlier	19,734	+/-920	15.3%	+/-0.7

Schuylkill County

	Estimate	Margin of Error	Percent	Margin of Error
Occupied housing units	**60,293**	**+/-912**	**100%**	**(X)**
Moved in 2005 or later	11,356	+/-794	18.8%	+/-1.2
Moved in 2000 to 2004	12,112	+/-788	20.1%	+/-1.2
Moved in 1990 to 1999	12,471	+/-721	20.7%	+/-1.2
Moved in 1980 to 1989	7,910	+/-591	13.1%	+/-1.0
Moved in 1970 to 1979	6,801	+/-510	11.3%	+/-0.8
Moved in 1969 or earlier	9,643	+/-579	16.0%	+/-1.0

Pennsylvania

	Estimate	Margin of Error	Percent	Margin of Error
Occupied housing units	**4,877,735**	**+/-6,828**	**99.90%**	**(X)**
Moved in 2005 or later	1,141,724	+/-8,691	23.4%	+/-0.2
Moved in 2000 to 2004	1,138,851	+/-7,875	23.3%	+/-0.2
Moved in 1990 to 1999	1,089,140	+/-7,880	22.3%	+/-0.2

Moved in 1980 to 1989	582,159	+/-6,089	11.9%	+/-0.1
Moved in 1970 to 1979	418,614	+/-4,504	8.6%	+/-0.1
Moved in 1969 or earlier	507,247	+/-4,438	10.4%	+/-0.1

Appendix C:

Initial Mailed Packet
Job History (Keep for Reference During Phone Survey)

To the best of your ability, please briefly tell us about **each job or occupation you held for at least 1 month since you left school (completed your education)**. Include full-time, seasonal work, part-time, volunteer work and military service (if you worked there at least 1 month). Also include your current job, even if you have had this for less than 1 month. Begin with your most recent job and continue back. Please estimate the time period if you cannot remember exact dates. *We will review this form during the phone survey*.

Job number	**Time period** **Mo/Yr to Mo/Yr**	**Job Title**	**Main Job Tasks**	**Company name (optional)**	**City and State of Workplace**
1	to				
2	to				
3	to				

4	to				
Job number	**Time period** **Mo/Yr to Mo/Yr**	**Job Title**	**Main Job Tasks**	**Company name (optional)**	**City and State of Workplace**
5	to				
6	to				
7	to				
8	to				

9	to				
10	to				
Job number	**Time period** **Mo/Yr to Mo/Yr**	**Job Title**	**Main Job Tasks**	**Company name (optional)**	**City and State of Workplace**
11	to				
12	to				

Task or Activity (place check-mark if present)	**In which of your jobs did you perform this task? List all *job numbers* that apply.**
Welding _	
Painting _	
Degreasing parts or equipment _	
Working with glues _	
Working with solvents or inks _	
Working where pesticides were used _ If yes: Purpose of pesticide? (check all that apply) To control plants/weeds (herbicide) __ To control insects (insecticide) __ To control rodents (rodenticide) __ Other (specify below) __	
Working with or near diesel-powered equipment _	
Working with or around live animals _	
Firefighting _	
Working with X-ray or radioactive material _	

For any of the jobs you listed, did you perform any of the following tasks or activities as part of your normal duties? Please check all that apply.

Appendix D :

Telephone script for to determine eligibility and complete questionnaire

Question 1: Hello my name is _____ and I am calling from the Geisinger Center for Health Research. May I please speak with

A) If Yes – Continue to **Question 2.**

B) If No – Ask for a better time to call back and set call back.

Question 2: I am calling to follow-up with you regarding the Drexel University Polycythemia Vera Research Study. Though we have not received your consent by mail, we are able to do the consent and questionnaire over the telephone now. May I proceed?

A) If Yes – Continue to **Question 3. [set callback if necessary]**

B) If No – SKIP to END REFUSAL

Question 3: Before beginning, I would like to give you some details about the study. First, can you confirm that you have continuously resided within the tri-county area of Carbon, Luzerne, or Schuylkill County between January 1, 2000 and December 31, 2008?

A) If Yes – Continue to **Question 4.**

B) If No – SKIP to END INELIGIBLE

Question 4: Thank you. You are being asked to take part in this research study because you are a resident of the tri-county region of Northeast Pennsylvania (Luzerne, Schuylkill, and Carbon counties). The purpose of this study is to investigate possible risk factors for certain blood diseases that have occurred among residents of Northeastern Pennsylvania.

- Study participants will include those who have certain blood diseases and also those who are not known to be affected by any blood disease.
- Taking part in this research study is voluntary. You may choose not to be in the study or withdraw from the study at any time.
- Appropriate measures will be made to keep your personal information confidential. You should know that according to Federal law the information that we are collecting from you is called Protected Health Information and that

you have certain rights regarding this information. We will protect the confidentiality of your protected health information in accordance with federal and state laws.

Do you have any questions about your privacy and confidentiality and your Protected Health Information?

A) If Yes: What questions do you have? (Answer questions). Skip to **Question 5**.

B) If No: Skip to **Question 5.**

Question 5: Your decision not to participate or to withdraw from the study will not involve any penalty or loss of benefits. It will not affect your access to health care at Geisinger Clinic or any other health care system.

Do you agree to participate in this research study by responding to survey questions and allowing us to use your Protected Health Information for this study?

If Yes: Great. We can now begin the survey, which will take between 45 minutes and an hour to complete [**set callback if necessary**]

If No: Skip to **END REFUSED.**

(Phone Questionnaire Complete)

Question 6: As an optional part of the study we would also like to offer a blood draw. We will send you an additional $25.00 Wal Mart gift card to thank you for your participation. By having the blood draw, you are giving the research team permission to collect, store, and test your blood samples (approximately 2 tablespoons) now or for future research studies to learn about, prevent, or treat characteristics related to the blood diseases under study. Would you be interested in the blood draw?

If Yes – Great! For the blood draw, we need a consent form mailed back to us in advance, so we will be mailing out a consent form for you to sign to agree to the blood draw. Skip to END

If No – Skip to END

END: Thank you again for your participation in the Drexel University Polycythemia Vera Research study. We truly appreciate your efforts. Should you have any questions or concerns please feel free to contact a member of the study team at 1-866-630-0798 option 2.

END REFUSED: I am sorry to hear that you are not interested in participating in the study. Thank you for your time. Have a nice day.

END INELIGIBLE: I am sorry that you do not meet the residency requirement for this study. If you'd like to speak with a member of the study team, please call 1-866-630-0798 option 2. Thank you for your time. Have a nice day.

Appendix E

Initial Mailed Packet
Residential History (Keep for Reference During Phone Survey)

To the best of your ability, please list **each primary residence you've lived in for 6 months or longer.** Begin with your current residence (#1) and work backwards, listing all residences back to age 21. Include your current residence even if you have lived there for less than 6 months. Fill in the address information to the best of your recollection, even if you don't have a complete address or ZIP Code. Please estimate the time period if you cannot remember exact dates. *We will review this form during the phone survey*.

For the residence description, please write the corresponding number in the space:
1) Single Family Home
2) Apartment
3) Condominium/Townhouse
4) Farm (please specify type: dairy, livestock, cash crops, etc.)
5) Mobile Home
6) Other (please specify)

Please use additional pages if needed.

For the residence description, please write the corresponding number in the space:

Residence number	Time period Mo/Yr to Mo/Yr	Street Address	City/Town	State and ZIP Code	How would you best describe this residence? (see key above)
1	to				
2	to				
3	to				

Residence number	Time period Mo/Yr to Mo/Yr	Street Address	City/Town	State and ZIP Code	How would you best describe this residence? (see key above)
4	to				
5	to				

1) Single Family Home
2) Apartment
3) Condominium/Townhouse
4) Farm (please specify: dairy, livestock, cash crops, or other)
5) Mobile Home

Residence number	**Time period** **Mo/Yr to Mo/Yr**	**Street Address**	**City/Town**	**State and ZIP Code**	**How would you best describe this residence? (see key above)**
6	to				
7	to				
8	to				
9	to				
10	to				

Vita

Ms. Gross-Davis received her MS from Drexel's College of Engineering and her BS in Biology from Cabrini College. She also serves as Assistant Professor in the Department of Environmental and Occupational Health at the School and is an Environmental Scientist with the Environmental Protection Agency (EPA) Region 3. In addition to her scientific training, Ms. Gross-Davis has 23 years of experience in the Federal Government as an environmental scientist for the EPA. Through her work at EPA she has experience in a broad range of Environmental Programs at the Federal, State and local government level focused on implementing and developing regulations and policy to support various initiatives, including: Community Air Toxics; Marcellus Shale air monitoring; Brownfield's Program and Hazardous Waste Program; Water Management Program and Partnerships and Innovation with Stakeholders. Part of her current work is to address urban air quality and the connection it has to health and our quality of life.