gastric intestinal metaplasia in transgenic mice. *Gastroenterology*, **122**, 689–696.
22. Mutoh,H., Hakamata,Y., Sato,K., Eda,A., Yanaka,I., Honda,S., Osawa,H., Kaneko,Y. and Sugano,K. (2002) Conversion of gastric mucosa to intestinal metaplasia in Cdx2-expressing transgenic mice. *Biochem. Biophys. Res. Commun.*, **294**, 470–479.
23. Laurén,P. (1965) The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. *Acta Pathol. Microbiol. Scand.*, **64**, 31–49.
24. Bai,Y.Q., Akiyama,Y., Nagasaki,H., Yagi,O.K., Kikuchi,Y., Saito,N., Takeshita,K., Iwai,T. and Yuasa,Y. (2000) Distinct expression of *CDX2* and GATA4/5, development-related genes, in human gastric cancer cell lines. *Mol. Carcinogen.*, **28**, 184–188.
25. Herman,J.G., Graff,J.R., Myohanen,S., Nelkin,B.D. and Baylin,S.B. (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl Acad. Sci. USA*, **93**, 9821–9826.
26. Herman,J.G., Umar,A., Polyak,K. *et al.* (1998) Incidence and functional consequences of *hMLH1* promoter hypermethylation in colorectal carcinoma. *Proc. Natl Acad. Sci. USA*, **95**, 6870–6875.
27. Akiyama,Y., Watkins,N., Suzuki,H. *et al.* (2003) *GATA-4* and *GATA-5* transcription factor genes and potential downstream antitumor target genes are epigenetically silenced in colorectal and gastric cancer. *Mol. Cell. Biol.*, **23**, 8429–8439.
28. Japanese Gastric Cancer Association. (1998) Japanese Classification of Gastric Carcinoma, 2nd English Ed. *Gastric Cancer*, **1**, 10–24.
29. Meulman,J.J. and Heiser,W.J. (2001) *SPSS Categories 11.0*. SPSS Inc., Chicago.
30. Inoshita,N., Yanagisawa,A., Arai,T., Kitagawa,T., Hirokawa,K. and Kato,Y. (1998) Pathological characteristics of gastric carcinomas in the very old. *Jpn. J. Cancer Res.*, **89**, 1087–1092.
31. Watabe,K., Nishi,M., Miyake,H. and Hirata,K. (1998) Lifestyle and gastric cancer: a case-control study. *Oncol. Rep.*, **5**, 1191–1194.
32. Ji,B.T., Chow,W.H., Yang,G., McLaughlin,J.K., Zheng,W., Shu,X.O., Jin,F., Gao,R.N., Gao,Y.T. and Fraumeni,J.F. Jr (1998) Dietary habits and stomach cancer in Shanghai, China. *Int. J. Cancer*, **76**, 659–664.
33. Cai,L., Zheng,Z.L. and Zhang,Z.F. (2003) Risk factors for the gastric cardia cancer: a case-control study in Fujian province. *World J. Gastroenterol.*, **9**, 214–218.
34. Yamane,T., Takahashi,T., Kuwata,K., Oya,K., Inagake,M., Kitao,Y., Suganuma,M. and Fujiki,H. (1995) Inhibition of N-methyl-N′-nitro-N-nitrosoguanidine-induced carcinogenesis by (−)-epigallocatechin gallate in the rat glandular stomach. *Cancer Res.*, **55**, 2081–2084.
35. Fujiki,H., Suganuma,M., Imai,K. and Nakachi,K. (2002) Green tea: cancer preventive beverage and/or drug. *Cancer Lett.*, **188**, 9–13.
36. Tajima,K. and Tominaga,S. (1985) Dietary habits and gastro-intestinal cancers: a comparative case-control study of stomach and large intestinal cancers in Nagoya, Japan. *Jpn. J. Cancer Res.*, **76**, 705–716.
37. Kono,S., Ikeda,M., Tokudome,S. and Kuratsune,M. (1988) A case-control study of gastric cancer and diet in northern Kyushu, Japan. *Jpn. J. Cancer Res.*, **79**, 1067–1074.
38. Kato,I., Tominaga,S., Ito,Y., Kobayashi,S., Yoshii,Y., Matsuura,A., Kameya,A. and Kano,T. (1990) A comparative case-control analysis of stomach cancer and atrophic gastritis. *Cancer Res.*, **50**, 6559–6564.
39. Imai,K., Suga,K. and Nakachi,K. (1997) Cancer-preventive effects of drinking green tea among a Japanese population. *Prev. Med.*, **26**, 769–775.
40. Fang,M.Z., Wang,Y., Ai,N., Hou,Z., Sun,Y., Lu,H., Welsh,W. and Yang,C.S. (2003) Tea polyphenol (−)-epigallocatechin-3-gallate inhibits DNA methyltransferase and reactivates methylation-silenced genes in cancer cell lines. *Cancer Res.*, **63**, 7563–7570.
41. Zhang,Y. and Talalay,P. (1994) Anticarcinogenic activities of organic isothiocyanates: chemistry and mechanisms. *Cancer Res.*, **54**, 1976–1981.
42. Murillo,G. and Mehta,R.G. (2001) Cruciferous vegetables and cancer prevention. *Nutr. Cancer*, **41**, 17–28.
43. Morse,M.A., Zu,H., Galati,A.J., Schmidt,C.J. and Stoner,G.D. (1993) Dose-related inhibition by dietary phenethyl isothiocyanate of esophageal tumorigenesis and DNA methylation induced by N-nitrosomethylbenzyl-amine in rats. *Cancer Lett.*, **72**, 103–110.
44. Fahey,J.W., Haristoy,X., Dolan,P.M., Kensler,T.W., Scholtus,I., Stephenson,K.K., Talalay,P. and Lozniewski,A. (2002) Sulforaphane inhibits extracellular, intracellular, and antibiotic-resistant strains of *Helicobacter pylori* and prevents benzo[a]pyrene-induced stomach tumors. *Proc. Natl Acad. Sci. USA*, **99**, 7610–7615.
45. Sipponen,P. and Correa,P. (2002) Delayed rise in incidence of gastric cancer in females results in unique sex ratio (M/F) pattern: etiologic hypothesis. *Gastric Cancer*, **5**, 213–219.
46. Imai,T. and Murayama,H. (1983) Time trend in the prevalence of intestinal metaplasia in Japan. *Cancer*, **52**, 353–361.

Received July 22, 2004; revised September 5, 2004; accepted September 29, 2004

Review Article

Perspectives on cancer immuno-epidemiology

Kei Nakachi, Tomonori Hayashi, Kazue Imai and Yoichiro Kusunoki

Department of Radiobiology/Molecular Epidemiology, Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami-ku, Hiroshima-shi, Hiroshima 732-0815

(Received August 19, 2004/Revised September 22, 2004/Accepted October 12, 2004)

Estimating human cancer risk based on host-environment interaction is one task of epidemiology, and it has provided indispensable knowledge for prevention of cancer. The recent development of gene-engineered mice has also provided solid evidence about the relationship between cancer development and immunity. The aim of this review is to discuss the possible contribution of epidemiology to understanding the role of immunity in host defense against cancer, and also to assess the involvement of inflammation in the occurrence of selected cancers. Here we look at the concepts of cancer immunosurveillance and infection-inflammation-cancer, and include a brief introduction to recent studies in humans and experimental animal models. It has been postulated for many years that the immune system has the ability to recognize and eliminate nascent transformed cells in the body (so-called cancer immunosurveillance hypothesis), and this idea has recently obtained strong support from animal experiments. In humans, follow-up studies among immunosuppressed transplant recipients revealed a remarkably increased risk of not only selected malignancies, but also cancers with no known viral etiology. On the other hand, a prospective cohort study among the general population revealed that individuals with low natural cytotoxic activity of peripheral blood lymphocytes had an increased risk of cancer development. More studies are warranted to allow the construction of a model for the interaction between host immunity, aging, and the environment. The host immune system is also involved in inflammatory responses to pathogen infection: insufficient immune function of the host, or repeated infection, may result in persistent inflammation, where growth/survival factors continuously act on initiated cells. The combined use of biomarkers will be necessary to define low-grade persistent inflammation in future cohort studies; and, in addition to these phenotype marker-based cohort studies, one plausible future direction will be a genomic approach that can be undertaken within cohort studies, looking at the genetic background underlying individual variations in phenotype markers. (Cancer Sci 2004; 95: 921–929)

Epidemiological studies investigate the association between cancer development and various environmental or/and host factors in human populations, providing models to estimate cancer risk as a quantitative function of these factors (e.g., exposure levels, physiological status) among individuals. We anticipate that epidemiological studies will work well under the following conditions: 1) the intensity of factors varies among individuals (being expected to produce substantial differences in cancer risk); 2) adequate measurements are available to evaluate the intensity or grade of factors (in the case of biomarkers); 3) a relevant basic biological concept or laboratory evidence-supported working hypothesis describing the relation between cancer and these factors is available; and 4) the association between cancer and these factors, if it exists, will contribute to cancer prevention. In this review, we discuss whether "cancer development and immunity" is a proper object of epidemiology from the above viewpoint.

The concept of multi-stage carcinogenesis implies that cancer prevention with different strategies at each stage is feasible. Recently, emphasis has been placed on defense mechanisms existing in different stages of carcinogenesis, such as detoxification of reactive metabolites derived from environmental carcinogens, trapping or decomposition of reactive oxygen species, DNA repair enzymes, and natural inhibitors of proliferating initiated cells.[1] The immune system may be the body's last line of defense against cancer development, and the concept of cancer immunosurveillance-routinely eliminating nascent transformed cells in the body-was first proposed by Burnet and Thomas.[2,3] However, despite accumulating evidence from *in vivo* studies that the immune system dominates the development of spontaneous tumors, observations in human populations have been limited, providing only marginal support for this concept. Since cancer immunosurveillance targets preclinically existing, nascent transformed cells, it is difficult to directly evaluate the immunological effects on cancer or pre-cancerous cells just emerging in the human body. Thus, epidemiological approaches such as long-term follow-up studies of human populations may be the most suitable way to assess the relation between host immunological status and future development of cancer in humans. Efficient epidemiological evaluation of host tumor immunity is thus different from efficient cancer immunotherapy, which targets clinically recognized cancer cells that are evading natural immunosurveillance and thereby acquiring a survival advantage.[4]

Another area where the host immune system is involved in cancer development may be in the sequential processes of infection-inflammation-cancer. Immunological features in the initial response to a pathogen in the host may in part determine how long and how strongly inflammation will continue after pathogen infection. The host will face chronic infection leading to persistent inflammation in the case of incomplete elimination of the corresponding pathogen, but, on the other hand, may retain homeostasis after successful eradication of the pathogen. Numerous observations of virus-related cancers have provided

E-mail: nakachi@rerf.or.jp

Abbreviations: STAT, signal transducer and activator of transcription; RAG, recombination activating gene; IFN, interferon; TNF, tumor necrosis factor; MHC, major histocompatibility complex; NK, natural killer; CTL, cytotoxic T lymphocyte; IL, interleukin; APC, antigen presenting cell; HLA, human leukocyte antigen; SNP, single nucleotide polymorphism; PGE, prostaglandin E. In Tables 1 and 2: WBC, white blood cell; ELISA, enzyme-linked immunosorbent assay; RT-PCR, reverse-transcriptase polymerase chain reaction; RANTES, regulation on activation, normal T cells expressed and secreted; MIG, monokine induced by interferon-γ; IP-10, interferon-γ-inducible protein-10; MCP-1, macrophage chemoattractant protein-1; CRP, C-reactive protein; ROS, reactive oxygen species; ESR, erythrocyte sedimentation rate; 8-OH-dG, 8-hydroxydeoxyguanosine; HPLC, high-performance liquid chromatography; ECD, electrochemical detector; MS, mass spectrometry; GC, gas chromatography; NICI, negative ion chemical ionization; HMdU, 5-hydroxymethyl-2'-deoxyuridine; Ig, immunoglobulin.

evidence that persistent inflammation involving repeated viral infection is a key step in carcinogenesis, although the immunological mechanisms underlying this process largely remain to be established.[5] Specifically, the environment-caused modification of host immune responses needs to be investigated in relation to cancer as well as other inflammation-related diseases: this might provide new and important insights into cancer prevention.

Cancer immunosurveillance

Cancer immunosurveillance may involve adaptive immune responses specific for antigens on malignant cells, as well as innate immune responses to non-self status or stress-induced ligands of transformed or malignant cells. Molecular changes that consistently occur in carcinogenesis of the cells may be recognized by the immune system as "flags" on target cells, and these aberrant molecules (neoantigens) may include: 1) products of oncogenes or tumor suppressor genes that are often mutated or products of other genes mutated due to genetic instability (e.g., Ras, Bcr/abl, p53),[6] 2) normal cellular proteins that are overexpressed or aberrantly expressed (e.g., MAGE, tyrosinase, gp100),[7] 3) oncogenic virus products (e.g., papillomavirus E6 and E7, EBNA-1, SV40 T antigen),[8] and 4) overexpression of stress-inducible proteins (e.g., NKG2D ligands: MICA, MICB, ULBPs).[9] Several mechanisms in which numerous other normal cellular molecules are involved can work to recognize, suppress, and/or eliminate tumor cells (Fig. 1). One of the key mechanisms in adaptive immunity for cases 1) to 3) involves the recognition of MHC/peptide complexes by T cells: tumor cells expressing mutated oncogene products can be eliminated *in vivo* by tumor-specific T cells that recognize MHC/peptide complexes in which the peptide components are encoded by mutant DNA sequences. However, some tumor cells can escape detection and survive when the mutated gene products in question are not presented as MHC/peptide complexes.[10]

On the other hand, innate immune responses for case 4) target a great variety of abnormal cells showing cellular transformation, infection, and distress, specifically in cases where the expression of MHC class I molecules is lost or downregulated ("missing-self"): NK cells can recognize and kill cells which overexpress the ligands of NKG2D, an activating NK receptor.[10] Here, NK cell effector functions are regulated by a balance between inhibitory receptors specific for MHC class I and activating receptors, although this NKG2D-mediating activation may be able to overcome the MHC class I-mediated inhibitory signaling in responding NK cells.[11] Clearly these two immunological mechanisms are complementary and work at different stages of the tumor-host interaction, providing as they do *in vivo* protection against the persistence of different types of tumor cells.

Granzymes, perforin, FasL and cytokines (such as IFN-γ) act as effector molecules for both T and NK cells to eliminate tumor cells; chemokines and their receptors are responsible for infiltration of lymphocytes into tumor tissue. In cases of infection by oncogenic viruses, such as hepatitis virus and HTLV-1, viral antigen presentation by HLA class I and II molecules to T cells, and subsequent T-cell mediated cytotoxicity and cytokine

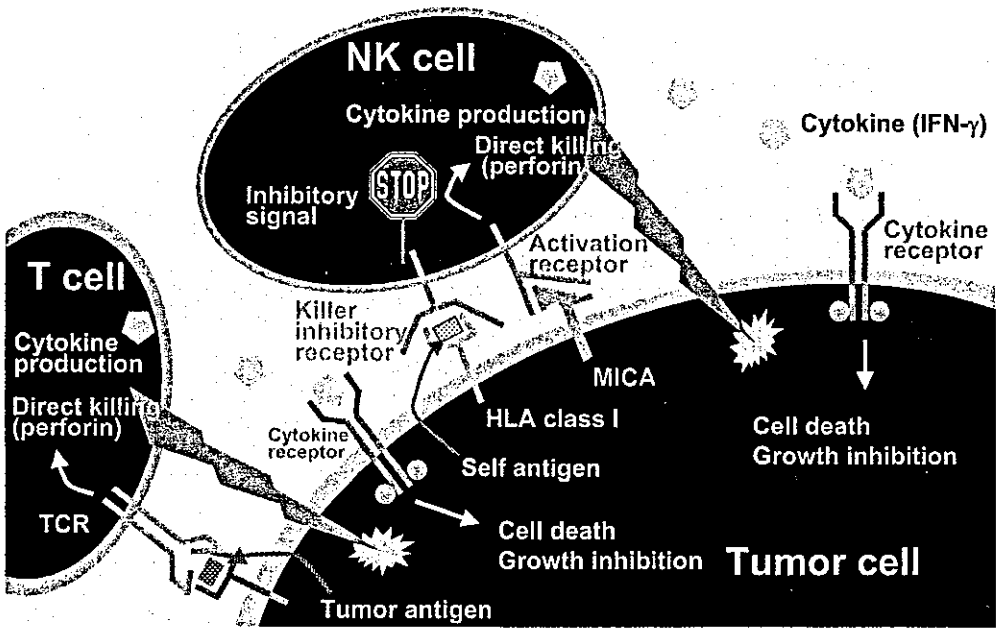


Fig. 1. Cells and molecules known to be involved in host immune responses to developing tumors. Tumor-specific cytotoxic T lymphocytes (CTL) recognize tumor antigens that are expressed in conjunction with HLA molecules and begin to directly kill tumor cells by secreting tumorcidal molecules (such as perforin) or to produce cytokines (such as IFN-γ) that suppress the growth of tumor cells. NK cell recognition is mediated by the opposing effects of two sets of NK receptors, activation and inhibitory receptors. Activation receptors recognize ligands (such as MICA) expressed on the target cell and transmit intracellular signals that initiate cytotoxicity; inhibitory receptors recognize cell-surface HLA class I molecules and generate counter-activating signals that block the induction of cytotoxicity. NK cell effector functions that kill or suppress tumor cells are almost identical to those of CTL. In the course of tumor progression, tumor cells tend to lose expression of HLA molecules and escape T cell recognition. Loss of HLA-class I expression (missing-self) on tumor cells engages NK cells to kill these cells.

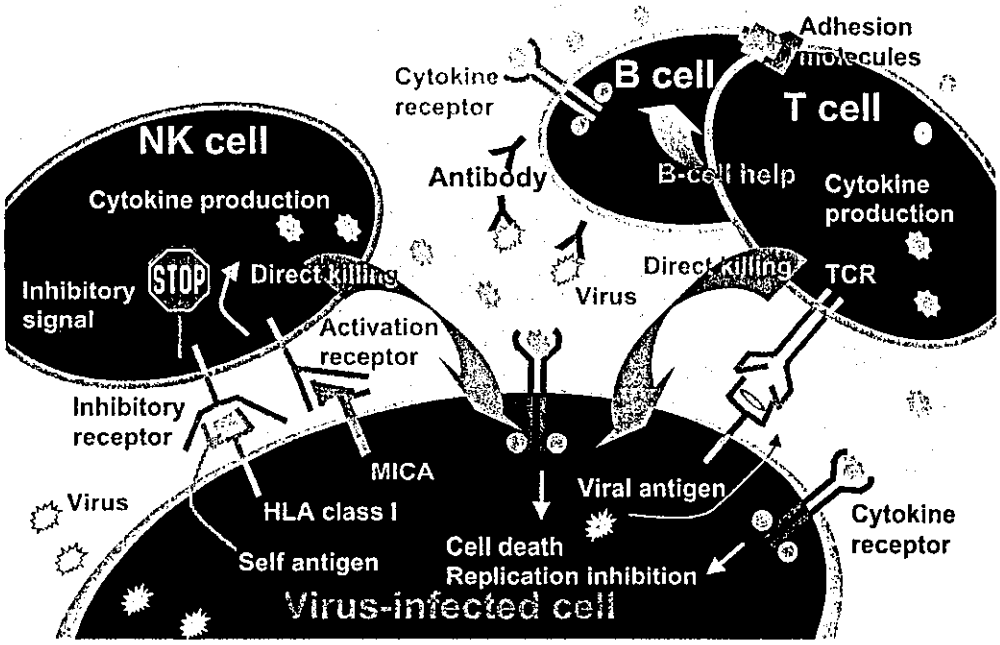


Fig. 2. In the case of immune responses to oncogenic virus infection, viral antigen presentation by HLA class I and II molecules to T cells initiates adaptive immune responses to the virus. Subsequent T-cell mediated cytotoxicity to virus-infected cells and T-cell cytokine production are key elements in control of the infection. NK cells can recognize virus-induced cellular antigens (such as MICA) and engage in eradication and suppression of virus-infected cells. T cells help B-cell production of virus-specific antibodies that can block viral replication by releasing cytokines and through direct cell-to-cell interaction.

production, are also key elements in the control of infection (Fig. 2). NK cells are also known to play important roles in eradication and suppression of virally infected cells.[10] Virus-specific antibodies that are produced by B cells with T-cell help may block infection of adjacent cells and thereby suppress viral replication. T and B cell interaction is mainly mediated by cytokines and cell adhesion molecules; and minor T-cell subsets, γδT and NKT cells, are also known to act as effector cells in the cancer elimination phase. Although the role of these immune effector mechanisms in tumor protection is not well defined, epidemiological approaches to investigating the association between cancer development and individual variations in the ability to mount these immune defense mechanisms are essential to establish the concept of cancer immunosurveillance and to develop a new basis for cancer prevention.

Observations in humans. One logical approach to examining the immunosurveillance hypothesis in humans is to determine whether patients with immunodeficiency, or immunosuppressed transplant recipients, show a greater incidence of cancer. A consistent finding in various follow-up studies of transplant recipients is a remarkably increased risk ratio (observed/expected ratio) of selected malignancies, many of which are associated with viruses such as Epstein-Barr virus (Hodgkin's disease), human papilloma virus (cervix cancer, anogenital cancer, and some skin cancers), human herpes virus 8 (Kaposi's sarcoma), and hepatitis B and C viruses (hepatocellular cancer).[12, 13] These observations have demonstrated that one relevant function of immunosurveillance is eradication of viruses, some of which may cause cancers, although it is still not clear whether the immune system can eliminate cancer cells with no known viral etiology. Recent studies have shown that transplant recipients have an increased risk of developing various cancers commonly observed in general populations, including those of the respiratory organs, digestive organs, and endocrine glands, which clearly demonstrates the role of immunological defense mechanisms in preventing the development of cancer.[12, 13] Of malignancies that develop in transplant recipients, the portion transmitted from donors is estimated to be less than 1%. In patients with various immunodeficiencies—such as Chediak-Higashi syndrome, X-linked lymphoproliferative syndrome, ataxia-telangiectasia, and the Wiskott-Aldrich syndrome—an increased incidence of selected cancers, such as non-Hodgkin's lymphoma, was observed; patients with adaptive immunodeficiency syndrome also showed 100-fold increase of Kaposi's sarcoma and non-Hodgkin's lymphoma.

However, these studies of immunodeficient populations have several limitations: 1) study subjects were relatively young and had therefore not reached the age when solid cancers are frequently seen (e.g., the mean age at transplantation was 43 years, and the mean age for diagnosis of malignancies was 48 years in the Cincinnati Transplant Tumor Registry) and the follow-up periods were short (in part, due to the patients' shortened lifespans and medical complications)[13]; 2) since immunodeficient people seem to carry widespread dysfunctions of the immune system, including both innate and adaptive immunity, it is difficult to assess the involvement of a specific immune function in cancer immunosurveillance, which also causes difficulty in extrapolating results obtained with immunodeficient people to the general population, who do not have obvious defects in the immune system and who have reached the "cancer-prone age." Since aging is the most important factor in the development of cancer, it is important to know how inter-individual differences in a particular immune function are associated with future development of common cancers among the

general population. In addition, the existence of pre-clinical cancer in the body may influence the immune function, so case-control studies seem to be inadequate for assessing the relation between cancer and immunological defense. Therefore, prospective cohort studies, using specific immunological biomarkers that are measurable with peripheral lymphocytes and stable during long periods of follow-up, are needed. Unfortunately, very few such studies are available.

In one prospective cohort study of the Japanese general population (the Saitama cohort study), an 11-year follow-up study recently revealed that individuals with medium and high natural cytotoxic activity of peripheral-blood lymphocytes—measured by the isotope-release method using K562 as target cells—had a reduced risk of developing cancer in all sites, whereas those with low cytotoxic activity had an increased cancer risk (Table 1).[14] This is the first evidence of the vital role played by natural immunological defense in the occurrence of common cancers among the general population who do not have obvious defects in their immune systems, indicating the possible feasibility of cancer immuno-prevention. Since natural cytotoxic activity is in part associated with selected lifestyle factors as well as mental stress, this cytotoxic activity will be a useful surrogate marker for future cancer prevention studies.[15,16] The findings also imply that individual variations in innate immune responses seen in the general population may generate large differences in cancer incidence with advancing years, specifically when people reach cancer-prone age. To date, though, no clear results have been obtained from studies using biomarkers of adaptive immunity. However, in one promising on-going cohort study, a subcohort of atomic-bomb survivors (the RERF immunological cohort study) has revealed a significant dose-dependent association between past experience of radiation exposure and attenuated immunity measured in terms of T-cell repertoire and functions, and cytokine levels, all of which are also associated with aging.[17–20] It is anticipated that a baseline measurement of various immunological markers of adaptive immunity in this unique cohort will answer some questions on cancer immunosurveillance and will provide a model for the interactions among host immunity, aging, and environment.

In addition, a genomic approach was recently undertaken in the Saitama and RERF cohort studies. To find genetic factors involved in individual variations of natural cytotoxic activity, age- and sex-matched Saitama cohort groups with high and low natural cytotoxic activity were compared in terms of *HLA class I* genotype frequencies: *B*1301, B*4403, B*5401, Cw*0401*, and *Cw*0702* were significantly associated with the activity (P=0.02, 0.02, 0.04, 0.03, and 0.004, respectively).[21] Specifically, *Cw*0702* is relatively frequent (11%) among the Japanese population. This phenotype-genotype association analysis within cohorts is now being extended to the genetic polymorphisms of NK cell receptors, a new genomic approach unique to cohort studies. In the RERF immunological cohort study, radiation effects on risk of type II diabetes were studied in terms of *HLA class II* haplotyping, indicating that individuals with a particular *HLA* haplotype, either *DQA1*03-DRB1*09* or *DQA1*0401-DRB1*08*, revealed an increased risk of type II diabetes dependently on their atomic-bomb radiation dose (trend P=0.0003).[22]

Table 1. Relative risk of cancer incidence for cytotoxic activity levels

	NK cell activity[1] (%)		
	Low	Medium	High
Men			
Age-adjusted	1.0	0.62 (0.38–1.03)[2]	0.72 (0.45–1.16)
Lifestyle-adjusted[3]	1.0	0.61 (0.37–1.02)	0.71 (0.44–1.16)
Women			
Age-adjusted	1.0	0.56 (0.31–1.01)	0.52 (0.28–0.95)
Lifestyle-adjusted	1.0	0.56 (0.31–1.04)	0.52 (0.29–0.98)
Both sexes			
Age-adjusted	1.0	0.59 (0.40–0.87)	0.63 (0.43–0.92)
Lifestyle-adjusted	1.0	0.60 (0.41–0.87)	0.64 (0.44–0.94)

1) Categorized by tertiles. Low: less than 42%, medium: 42–58%, high: more than 58% for men; low: less than 34%, medium: 34–51%, high: more than 51% for women.
2) 95% confidence interval.
3) Adjusted for age, relative body weight, cigarette smoking, alcohol consumption, and intake of green vegetables.

Experimental animal studies. The cancer immunosurveillance hypothesis has been tested using numerous immunocompromised animals in which spontaneous and/or carcinogen-induced tumor development was assessed. Several lines of experiments using athymic nude mice found no significant increase in tumor formation as compared with euthymic immunocompetent mice,[23] and these negative results initially gave some tumor immunologists an unfavorable view of this hypothesis. However, as modern immunology has begun to explain abnormalities in the immune system in terms of deficiencies of particular genes, various gene-knockout mice have become available for testing the immunosurveillance hypothesis. Mice deficient in one of several key molecules (IFN-γ, IFNGRI, and STAT-1) involved in the IFN-γ system more frequently developed spontaneous and/or carcinogen-induced tumors than did wild-type mice.[24,25] *Rag2* gene ablation, which results in lack of lymphocytes mediating adaptive immunity, also appeared to increase susceptibility to spontaneous and/or carcinogen-induced cancers.[24] Interestingly, mice deficient in both RAG2 and STAT-1 did not differ in overall incidence of tumors from those deficient in only one, suggesting that the IFN-γ system may be a major effector mechanism for tumor suppression through adaptive immunity.[24]

Another key effector molecule for immunological tumor control has been identified from studies with perforin-knockout mice, which also show increased susceptibility to tumor development[25,26]: perforin is a component of cytolytic granules of CTL and NK cells, and mice deficient in both perforin and IFN-γ showed a small increase in tumor induction compared with those lacking only one of the two immune mediators, suggesting the existence of cross-talk between innate and adaptive immunity for resisting tumor formation.[25] IL-12 is a potent inducer of Th1, which produces IFN-γ and exerts anti-tumor immunity by activating both CTL and NK cells, and mice defective in one of the IL-12 subunits are also known to be more susceptible to chemical carcinogenesis.[27] This anti-tumor cytokine is produced by macrophages and dendritic cells and plays a key role in the transition from innate to adaptive immunity, again suggesting cross-talk between these immune systems in cancer immunosurveillance.

Another important step in experimental animal studies on cancer immunosurveillance was demonstrating the possible involvement of NK-mediated effector mechanisms in the suppression of tumor formation. Previous observations with nude mice,[23] which challenged the cancer immunosurveillance hypothesis, ignored the fact that nude mice have a potential innate immune system including NK cell function. NK-deficient beige mice, which have a defect in cytolytic granule formation that also affects CTL and macrophages, have an increased incidence of spontaneous and induced primary oncogenesis.[28] In addition, antibody depletion studies using anti-NK1.1 or anti-asialo-GM1 antibody, which can deplete NK cells as well as NKT cells or

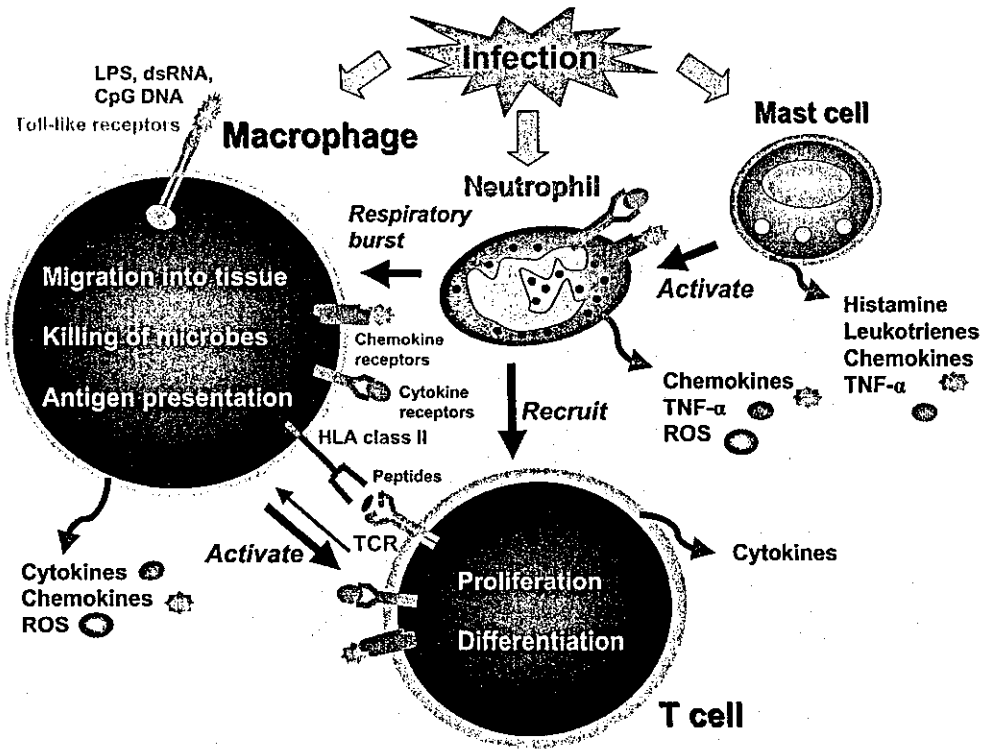


Fig. 3. Flow of inflammatory processes. 1) "Danger signals" from tissue trauma. The bioactive peptides released from neurons in response to pain activate mast cells, the intracellular proteins released from destroyed cells activate macrophages, and pathogen-associated patterns also activate macrophages through toll-like receptors. 2) Mast cells, the first responder, release histamine, luekotrienes, TNF-α, other cytokines/chemokines, and tryptases. 3) Neutrophils are activated by TNF-α and leukotrienes produced by mast cells; activation of matrix metalloprotenases; tissue breakdown. 4) Macrophages are activated by pathogen-associated patterns; macrophage-derived TNF-α and chemokines activate more neutrophils and recruit lymphocytes, in conjunction with PGE2, from mast cells and defensins from neutrophils. 5) Inflammatory responses (activation of mast cells and neutrophils) evolve into immune responses, i.e., activation of macrophages, dendritic cells (DCs), T cells and B cells.

activated macrophages, revealed that NK cells are important in preventing tumor induction by a chemical carcinogen.[27] Although determining the specific roles of NK cells in cancer immunosurveillance is hampered by the lack of mouse models completely defective in NK cells but normal in T and B cells, it is likely that NK cells participate in various stages of tumor immunity, including cancer immunosurveillance, as has been indicated by a follow-up study in humans. NK cell recognition is mediated by the opposing effects of two different types of NK receptors, activation and inhibitory receptors: activation receptors recognize stress-induced ligands that are expressed on the target cell, and then transmit intracellular signals that initiate cytotoxicity; inhibitory receptors recognize cell-surface MHC class I molecules and generate counter-activating signals that block the induction of cytotoxicity. Besides being a vital player in innate immunity, NK cells influence subsequent adaptive immune responses by releasing cytokines and chemokines that induce growth and differentiation of various immune cells.

Infection, inflammation, and cancer

Cell proliferation does not by itself cause cancer: growth/survival factors enriched at sites of inflammation specifically promote the proliferation of initiated cells. Once tissue trauma has healed, the inflammation associated with cell proliferation (required for tissue-regeneration) ends. However, when inflammation is sustained and becomes chronic, continuous growth stimuli work on initiated cells and reactive oxygen species cause genotoxic damage, generating dysplastic changes (atypical cells). Therefore, for cancer cell development at sites of inflammation, it is critical whether the inflammation becomes persistent or not.

Inflammation is a sequential process of responding to the trauma often caused by microbial infection; in the process, various soluble factors and infiltrating or recruited cells (such as lymphocytes and leukocytes) become involved while interacting with each other in several steps: 1) recognition of tissue penetration by pathogens or tissue injury; 2) beckoning, instruction, and dispatch of cells (infiltration of lymphocytes); 3) eradication of pathogens and killing of infected cells; 4) liquefaction of surrounding tissue to prevent microbial metastasis; 5) healing of damaged tissue. Throughout, several checkpoint signals determine the advance or standstill of inflammatory response; if this sequential process is hindered at any step, or if repeated infections occur within the host, the inflammatory process may become stalled, resulting in persistent inflammation.[29]

Innate immune responses are induced by pathogen-associated patterns (e.g., lipopolysaccharide (LPS), double-stranded (ds) RNA, CpG DNA), which are recognized by toll-like receptors on macrophages,[30] and/or by the NKG2D and other NKp receptor ligands expressed on infected cells.[31] The cytokine cascade plays an important role in augmentation and suppression of immune response to pathogens: cytokines released from APC or T cells, such as IL-6, TNF-α, and IFN-γ act as effector molecules in inflammation induced by microbial infection; chemokines and their receptors are then involved in migration of immune cells into inflammation sites (Fig. 3).[5,29] Hence, insufficient immune functions of the host may result in persistent inflammation because of failure to completely eradicate pathogens or infected cells, resulting in repeated destruction and regeneration of tissue. When initiated cells at sites of persistent inflammation continue to proliferate—interacting with inflam-

matory cells and growth factors that specifically act on the initiated cells (e.g., TNF-α)—the inflammatory process becomes a crucial step in carcinogenesis. In fact, many cancers are thought to be associated with inflammation caused by immunologically uncontrolled infections: colon carcinogenesis arising in individuals with inflammatory bowel diseases, chronic ulcerative colitis and Crohn's disease, esophageal carcinoma with reflux esophagitis, gastric cancer with atrophic gastritis, liver cancer with hepatitis, lung cancer with interstitial pneumonia, etc.

Observations in humans. It is said that 18% of cancer cases worldwide can be ascribed to infections with various pathogens, which include *Helicobacter pylori* (gastric cancer, 490,000 cases a year), human papillomavirus (cancer of the cervix and other sites, 550,000), hepatitis B and C viruses (hepatocellular carcinoma, 390,000), Epstein-Barr virus (lymphomas and nasopharyngeal cancer, 99,000), human herpes virus 8 (Kaposi's sarcoma, 54,000), *Schistosoma haematobium* (bladder cancer, 9000), human T-cell lymphotropic virus (adult T-cell leukemia, 2700), and *Opisthorchis viverrini* (cholangiocarcinoma, 800).[32] For some of these cancers, the immunogenetic status of the host has been investigated in terms of HLA typing and SNPs in cytokine genes.[33] Specifically, the identification of HLA class II types that are sensitive or resistant to human T-cell lymphotropic virus has demonstrated a role of host immunity in virus-associated carcinogenesis.

The oncogenic processes in virus-related cancer are greatly influenced by a series of immune effector mechanisms: virus-infected cells that have encountered the immune system eventually go through processes involving escape from immunological recognition and cytolysis, and the cell transformation that accompanies rapid proliferation causes frequent gene mutations. How an individual's defense system undertakes these processes is thought to depend on individual ability to mount immunity in response to infection with a particular virus. Decreased immunity to infection with such a virus when complete elimination of the extrinsic antigen has failed may be closely related to carcinogenesis that results from continuous inflammation, and repeated destruction and regeneration of tissue, causing mutations. Among many cancers in which inflammation is considered to be involved, some may also be associated with production of carcinogenic proteins by infected microbes, e.g., oncoprotein CagA by *H. pylori* in gastric cancer, oncoproteins X by hepatitis B virus (HBV) and core protein by hepatitis C virus (HCV) in hepatocellular carcinoma, oncoprotein E6/7 by human papilloma virus (HPV) in cervical cancer, and oncoprotein Tax by human T-cell lymphoma virus (HTLV-I) in adult T-cell leukemia. HLA molecules play an important role in the recognition of antigens derived from carcinogenic proteins that have the potential to transform cells infected with these microbes, possibly ensuring surveillance of transformed cells.[34] In some cases, a particular HLA class II molecule may lack the capacity for binding to the peptide anchor motif needed to recognize an oncoprotein, and thereby fail to induce CTL responses to transforming cells and allow generation of a specific type of cancer. In support of this notion, there are numerous reports suggesting an association between susceptibility to cancer and *HLA class II* genotype.[33]

Apart from infection-related cancers, many cancers have been associated with persistent inflammation: lung cancer associated with asbestosis or silicosis; colon cancer with inflammatory bowel disease, Crohn's disease, and chronic ulcerative colitis; pancreas cancer with chronic pancreatitis; esophageal cancer with reflux esophagitis or Barrett's esophagus; MALT lymphoma with Sjögren syndrome; melanoma with UV-caused skin inflammation; bladder cancer with chronic cystitis or bladder inflammation; oral squamous cell carcinoma with gingivitis.[5] These findings may imply that persistent inflammation itself has carcinogenic activity, due to production of reactive oxygen species, tumor promotion activity of inflammatory cytokines, and induction of genetic instability. One recent cohort study revealed that plasma levels of C-reactive protein were an excellent predictor of the risk of colon cancer, demonstrating that subclinical persistent inflammation may underlie colon carcinogenesis in general.[35] Interestingly, C-reactive protein levels were unchanged by administration of nonsteroidal anti-inflammatory drugs.

One problem in designing epidemiological studies to examine the relation between low-grade inflammation and cancer seems to be adequate selection of biomarkers that can define low-grade, persistent inflammation. It may be desirable to use a combination of inflammation-related markers such as plasma levels of IL-6, IL-10, TNF-α, and IFN-γ, along with C-reactive protein (CRP), together with erythrocyte sedimentation rate (ESR), whose validity has been demonstrated in the cohort of atomic-bomb survivors.[20] Since all these biomarkers are closely related to aging, the effect of aging and environmental factors on inflammatory status can be investigated in relation to occurrence of aging-related diseases, such as cancer. Candidate biomarkers which have been or could be used in studies of immuno-epidemiology are listed in Table 2. Environmental factors or events that potentially influence the immunological/inflammatory status of the host should be identified, and their relationship to the incidence of cancer should be intensively investigated. In addition to various pathogens, exposure to chemical carcinogens and radiation may induce impairments in the immune system on some occasions, resulting in low-grade chronic inflammation and eventually leading to enhanced risk of cancer development.

One long-term prospective cohort study has examined the effects of radiation on the health of atomic-bomb survivors. To our surprise, even now, more than 50 years after the bombings, impairments in T-cell immunity are radiation-dose-dependently observed among a sub-cohort of atomic-bomb survivors, along with increased levels of plasma inflammatory cytokines and other inflammation markers.[17, 18, 20] In fact, atomic-bomb survivors even today continue to suffer from increased risk of cancer, cardiovascular disease, and hepatitis. These late effects pose serious, as yet unanswered, questions about the mechanisms involved. We hypothesize that T-cell impairments caused by radiation may generate age-associated chronic low-grade inflammation, which may in part be responsible for increased risk of diseases among atomic-bomb survivors. Decreased CD4 helper T-cell counts of the survivors appeared to be significantly associated with increased levels of IL-6 and CRP.[36] We found that both radiation exposure and increased age were associated with increases in selected plasma inflammatory biomarkers (Table 3), indicating that the effect of radiation could be further estimated in terms of acceleration of aging.[20, 36] Among the inflammatory biomarkers we examined for the effects of increased age and radiation dose, the increased levels of TNF-α, IL-10, IL-6, ESR, CRP, and IgA per Gy corresponded, on average, to an increase in age of 10 years (range, 5 to 15); atomic-bomb survivors' average radiation dose was 0.2 Gy, corresponding to about 2 years (range, 1 to 3) of aging. This may provide a hint as to why the incidence of cancer and some inflammation-associated diseases among atomic-bomb survivors remains high even when so much time has elapsed, as well as a model for understanding the effects of various environmental factors on aging-related diseases in general. This cohort study clearly shows the significance of repeated clinical examination, measurement of various immunological markers (some of which are listed in Table 2), and preservation of biological materials, through more than 50 years of follow-up.

Experimental animal studies. Although numerous factors and cells are involved in the complicated process of inflammation, cytokines are assumed to play a key role in the crossover of in-

Table 2. Candidate biomarkers for immuno-epidemiology

Phenotype	Marker(s)	Function(s)	Method	Reference(s)
Cell numbers				
	WBC	Inflammation	Cell counting	46
	Neutrophils	Innate immunity/inflammation	Cell counting	46
	CD4	Cellular immunity/helper T cell	Flow cytometry	47
	CD8	Cellular immunity/cytotoxic T cell	Flow cytometry	47
	CD19	Humoral immunity/B cell	Flow cytometry	47
	CD16/CD56	Innate immunity/NK cell	Flow cytometry	47
	CD45RA/CD45RO	Naïve/memory T cell	Flow cytometry	47
	Th1	Cellular immunity/helper T cell	Intracellular staining of IFN-γ	48
	Th2	Cellular immunity/helper T cell	Intracellular staining of IL-4	48
Cell activities				
	NK activity	Innate immunity	Isotope release	14
	T cell proliferation	Blast formation of T cell by mitogen	[^{3}H]thymidine incorporation	49
Cytokines				
	IL-6	Pro-inflammation	ELISA, Real time RT-PCR	20
	IL-8	Pro-inflammation	ELISA, Real time RT-PCR	50
	TNF-α	Pro-inflammation	ELISA, Real time RT-PCR	20, 50
	IL-1β	Pro-inflammation	ELISA, Real time RT-PCR	50
	IL-10	Anti-inflammation	ELISA, Real time RT-PCR	20, 50
	IFN-γ	Pro-inflammation	ELISA, Real time RT-PCR	20, 50
Chemokines				
	RANTES	Inflammation/recruitment of lymphocytes	ELISA, Real time RT-PCR	51
	MIG	Inflammation/recruitment of lymphocytes	ELISA, Real time RT-PCR	52
	IP-10	Inflammation/recruitment of lymphocytes	ELISA, Real time RT-PCR	52
	MCP-1	Inflammation/recruitment of lymphocytes	ELISA, Real time RT-PCR	53
Plasma/serum inflammatory markers				
	CRP	Inflammation	ELISA	35
	Metabolites of ROS	Inflammation/ROS production	Total ROS assay system	54
	ESR	Inflammation	Wintrobe method	46
	Sialic acid	Inflammation	Enzyme assay	46
	Haptoglobin	Inflammation	Nephrometry	46
	HMdU	Inflammation/DNA damage	ELISA	55
Tissue/cell inflammatory marker				
	8-OH-dG	Inflammation/DNA damage	HPLC/ECD	56
	Etheno DNA adduct	Inflammation/DNA damage	HPLC/MS, GC/MS, GC/NICI/MS	57

Table 3. Multivariate model of the effects of age, sex, and radiation dose on inflammatory biomarkers and immunoglobulins[1]

Variable	Percent increments (95% confidence intervals)										
	TNF-α	IFN-γ	IL-10	IL-6	CRP	ESR	Total Igs	IgG	IgA	IgM	IgE
Age per 10 years	15 (9, 20)	4 (−4, 12)	8 (4, 13)	24 (19, 30)	25 (13, 38)	15 (9, 20)	3 (1, 6)	3 (1, 6)	5 (2, 9)	−6 (−11, 14)	2 (−11, 14)
Female sex[2]	15 (2, 30)	−8 (−23, 10)	6 (0, 12)	8 (−41, 18)	0 (−25, 33)	17 (9, 24)	5 (0, 10)	7 (1, 13)	−9 (−17, −1)	14 (1, 28)	−51 (−63, −34)
Radiation dose per Gy	7 (1, 15)	12 (2, 23)	6 (0, 12)	13 (6, 20)	39 (20, 62)	17 (9, 24)	3 (1, 6)	2 (−1, 5)	8 (3, 13)	9 (2, 15)	14 (−3, 32)
Estimated aging by radiation (years per Gy)[3]	5 (0, 10)	29 (−29, 88)	6 (−1, 14)	5 (2, 8)	14 (4, 24)	11 (5, 17)	12 (−1, 26)	6 (−4, 17)	15 (1, 29)	−14 (−29, 2)	90 (−682, 861)

1) Subjects were a total of 442 atomic-bomb survivors who did not have a history of cancer or inflammatory-associated diseases (e.g., current cold, chronic bronchitis, collagen disease, arthritis, myocardial infarction).
2) Percentage change, female versus male.
3) Estimated by the δ-method.

flammation and cancer. Development of cytokine-gene knockout mice has demonstrated the vital role of pro-inflammatory cytokines in carcinogenesis: TNF-α-deficient mice developed a significantly smaller number of tumors than did wild-type mice in two-stage skin carcinogenesis experiments, demonstrating that TNF-α is the key cytokine by which inflammation acts as a tumor promoter.[37, 38] The IL-1 knockout mouse model implies that host-derived IL-1α and IL-1β are required for control of tumor angiogenesis and invasiveness in a melanoma model.[39] In a urethane carcinogenesis experiment, TNF-α and IL-10 deficiencies showed contrasting effects on lung tumor susceptibility,[40] and the pro-inflammatory cytokines, TNF-α, IL-1, and IL-6, seem to play different roles in tumor promotion and cell transformation.[41] In addition to these cytokines, macrophage

migration inhibitory factor (MIF) has been reported to amplify carcinogenic DNA damage by suppressing the transcriptional activity of p53 and by-passing p53 regulatory functions.[42] We thus anticipate that a network of inflammatory signals, with discrete roles of cytokines/chemokines and their interactions, will be intensively studied in relation to cancer development.

Macrophages sense a variety of microbes through toll-like receptors that recognize pathogen-associated patterns, while NK cells recognize host proteins expressed after infection through NK-activating receptors, such as NKG2D. Recently, the interaction between innate/adaptive immunity and inflammatory response has been delineated: murine macrophages, which are activated with LPS through toll-like receptor, express ligands (RAE-1) that are recognized by NKG2D receptor on NK cells, thus implying a mechanism by which NK cells and infected macrophages directly interact during an innate immune response to infection.[43] With HBV transgenic mice, CTL-mediated destruction of infected hepatocytes reportedly induces long-lasting hepatocellular regeneration, oxidative DNA damage, and clonal expansion, eventually resulting in hepatocellular carcinoma.[44] This study leads to the quintessential question: are pathogen-specific functions essentially required for cancer development, in addition to persistent inflammation itself (including induction of inflammatory cytokines)?

Conclusions and perspectives

It is anticipated that cancer epidemiology will eventually clarify the roles of immunity in protecting the host from nascent transformed cells and in regulating inflammatory responses to pathogens. Although the recent development of gene-engineered mice has provided solid evidence for cancer immunosurveillance and for the inflammation-cancer sequence, reliable estimation of cancer risk for individually differing immunological competence can be performed only in epidemiology, which could also identify high-risk individuals and aim at cancer prevention based on immunological up-regulation. One advantage of immuno-epidemiology may be the array of biomarkers listed in Table 2, which demonstrates that peripheral blood can reflect the systemic status of host immunity. On the other hand, the fact that the immune system is easily influenced by the existence of cancer in the body, even when it is in a preclinical stage, narrows down the study methods to prospective cohort studies. Although only a few such cohort studies are available at present (e.g., the Saitama cohort study and the RERF immunological cohort study), these studies should be expanded and extended in the future to answer the numerous questions concerning the roles of immune cells in cancer surveillance and inflammation, the characteristics of inflammation associated with cancer development, the effects of environment/lifestyle factors on the immune system, and the interaction between aging and immunity in the occurrence of cancer and other diseases.

Another important issue to be considered is the genetic background underlying individual variations in immune and inflammatory responses. *HLA* haplotyping has been intensively studied in relation to cancer among different races, and genetic polymorphisms of various cytokines and their receptors have also been investigated, mostly in case-control study design. One representative study is a large-scale case-control study which revealed that genetic polymorphisms of inflammatory cytokines including *IL-1β* influenced the risk of gastric cancer by modulating the pH of gastric juice and the growth environment of *Helicobacter pylori*.[45] A possible advantage of this genomic approach is that the involvement of immune-related genes can be readily examined in case-control studies, although any mechanistic interpretation (or conclusion on the functional significance of particular genetic polymorphisms considered in studies) must be made separately. However, risk estimation in these studies is made for particular polymorphisms of genes, not for the function or role of the genes.

On the other hand, cohort studies seem to have an advantage over case-control studies for the genomic approach: genomic analysis comparing cohort members with high and low values of particular phenotype markers can readily be performed, along with follow-up studies that reveal the association between these markers and cancer development. This phenotype-genotype association analysis within cohort studies may clarify the genetic background of those phenotype markers that are directly related to cancer risk, and possibly lead to their use as surrogate biomarkers for cancer prevention.

The work performed in our laboratory was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and for Cancer Research from the Ministry of Health, Labour and Welfare, Japan.

1. Jakóbisiak M, Lasek W, Golab J. Natural mechanisms protecting against cancer. *Immunol Lett* 2003; **90**: 103–22.
2. Burnet F. Cancer—a biological approach. *Br Med J* 1957; **1**: 841–7.
3. Thomas L. In: Lawrence HS, editor. Cellular and humoral aspects of the hypersensitive status. New York: Hoeber-Harper; 1959. p. 529–32.
4. Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoediting: from immunosurveillance to tumor escape. *Nat Immunol* 2002; **3**: 991–8.
5. Coussens LM, Werb Z. Inflammation and cancer. *Nature* 2002; **420**: 860–7.
6. Cheever MA, Disis ML, Bernhard H, Gralow JR, Hand SL, Huseby ES, Qin HL, Takahashi M, Chen W. Immunity to oncogenic proteins. *Immunol Rev* 1995; **145**: 33–59.
7. Boon T, Coulie PG, Van den Eynde B. Tumor antigens recognized by T cells. *Immunol Today* 1997; **18**: 267–8.
8. Ostrand-Rosenberg S. Animal models of tumor immunity, immunotherapy and cancer vaccines. *Curr Opin Immunol* 2004; **16**: 143–50.
9. Raulet DH. Roles of the NKG2D immunoreceptor and its ligands. *Nat Rev Immunol* 2003; **3**: 781–90.
10. Wu J, Lanier LL. Natural killer cells and cancer. *Adv Cancer Res* 2003; **90**: 127–56.
11. Cerwenka A, Baron J, Lanier L. Ectopic expression of retinoic acid early inducible-1 gene (RAE-1) permits natural killer cell-mediated rejection of a MHC class I-bearing tumor *in vivo*. *Proc Natl Acad Sci USA* 2001; **98**: 11521–6.
12. Sheil AG, Disney AP, Mathew TH, Livingston BE, Keogh AM. Lymphoma incidence, cyclosporine, and the evolution and major impact of malignancy following organ transplantation. *Transplant Proc* 1997; **29**: 825–7.
13. Penn I. Posttransplant malignancies. *Transplant Proc* 1999; **31**: 1260–2.
14. Imai K, Matsuyama S, Miyake S, Suga K, Nakachi K. Natural cytotoxic activity of peripheral-blood lymphocytes and cancer incidence: an 11-year follow-up study of a general population. *Lancet* 2000; **356**: 1795–9.
15. Nakachi K, Imai K. Environmental and physiological influences on human natural killer cell activity in relation to good health practices. *Jpn J Cancer Res* 1992; **83**: 798–805.
16. Imai K, Nakachi K. Personality types, lifestyle, and sensitivity to mental stress in association with NK activity. *Int J Hyg Environ Health* 2001; **204**: 67–73.
17. Kusunoki Y, Yamaoka M, Kasagi F, Hayashi T, Koyama K, Kodama K, MacPhee DG, Kyoizumi S. T cells of atomic bomb survivors respond poorly to stimulation by *Staphylococcus aureus* toxins *in vitro*: does this stem from their peripheral lymphocyte populations having a diminished naive CD4 T-cell content? *Radiat Res* 2002; **158**: 715–24.
18. Yamaoka M, Kusunoki Y, Kasagi F, Hayashi T, Nakachi K, Kyoizumi S. Decreases in percentages of naive CD4 and CD8 T cells and increases in percentages of memory CD8 T-cell subsets in the peripheral blood lymphocyte populations of A-bomb survivors. *Radiat Res* 2004; **161**: 290–8.
19. Kusunoki Y, Yamaoka M, Kasagi F, Hayashi T, MacPhee DG, Kyoizumi S. Long-lasting changes in the T-cell receptor V beta repertoires of CD4 memory T-cell populations in the peripheral blood of radiation-exposed people. *Br J Haematol* 2003; **122**: 975–84.
20. Hayashi T, Morishita Y, Kubo Y, Kusunoki Y, Hayashi I, Kasagi F, Hakoda M, Kyoizumi S, Nakachi K. Long-term effects of radiation dose on inflammatory markers in atomic bomb survivors. *Am J Med* 2005; in press.
21. Hayashi T, Imai K, Kusunoki Y, Hayashi I, Kyoizumi S, Tahara E, Nakachi K. HLA genotyping is involved in inter-individual variations of NK activity.

In: Skamene E, editor. Immunology2004: Medimond S.r.l; 2004. p. 21–25.
22. Hayashi T, Fujiwara S, Morishita Y, Kusunoki Y, Nakashima E, Nakanishi S, Suzuki G, Nakachi K, Kyoizumi S. HLA haplotype is associated with diabetes among atomic bomb survivors. *Hum Immunol* 2003; **64**: 910–6.
23. Stutman O. Chemical carcinogenesis in nude mice: comparison between nude mice from homozygous matings and heterozygous matings and effect of age and carcinogen dose. *J Natl Cancer Inst* 1979; **62**: 353–8.
24. Shankaran V, Ikeda H, Bruce AT, White JM, Swanson PE, Old LJ, Schreiber RD. IFNgamma and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature* 2001; **410**: 1107–11.
25. Street SE, Trapani JA, MacGregor D, Smyth MJ. Suppression of lymphoma and epithelial malignancies effected by interferon gamma. *J Exp Med* 2002; **196**: 129–34.
26. Street SE, Cretney E, Smyth MJ. Perforin and interferon-gamma activities independently control tumor initiation, growth, and metastasis. *Blood* 2001; **97**: 192–7.
27. Smyth MJ, Thia KY, Street SE, Cretney E, Trapani JA, Taniguchi M, Kawano T, Pelikan SB, Crowe NY, Godfrey DI. Differential tumor surveillance by natural killer (NK) and NKT cells. *J Exp Med* 2000; **191**: 661–8.
28. Haliotis T, Ball JK, Dexter D, Roder JC. Spontaneous and induced primary oncogenesis in natural killer (NK)-cell-deficient beige mutant mice. *Int J Cancer* 1985; **35**: 505–13.
29. Nathan C. Points of control in inflammation. *Nature* 2002; **420**: 846–52.
30. Akira S, Takeda K. Toll-like receptor signalling. *Nat Rev Immunol* 2004; **4**: 499–511.
31. Cerwenka A, Lanier LL. NKG2D ligands: unconventional MHC class I-like molecules exploited by viruses and cancer. *Tissue Antigens* 2003; **61**: 335–43.
32. WHO, IARC. World cancer report. In: Stewart BWKP, editor. Lyon: IARC-Press; 2003.
33. Little AM, Stern PL. Does HLA type predispose some individuals to cancer? *Mol Med Today* 1999; **5**: 337–42.
34. Yashiki S, Fujiyoshi T, Arima N, Osame M, Yoshinaga M, Nagata Y, Tara M, Nomura K, Utsunomiya A, Hanada S, Tajima K, Sonoda S. HLA-A*26, HLA-B*4002, HLA-B*4006, and HLA-B*4801 alleles predispose to adult T cell leukemia: the limited recognition of HTLV type I tax peptide anchor motifs and epitopes to generate anti-HTLV type I tax CD8(+) cytotoxic T lymphocytes. *AIDS Res Hum Retroviruses* 2001; **17**: 1047–61.
35. Erlinger TP, Platz EA, Rifai N, Helzlsouer KJ. C-Reactive protein and the risk of incident colorectal cancer. *JAMA* 2004; **291**: 585–90.
36. Hayashi T, Kusunoki Y, Hakoda M, Morishita Y, Kubo Y, Maki M, Kasagi F, Kodama K, Macphee DG, Kyoizumi S. Radiation dose-dependent increases in inflammatory response markers in A-bomb survivors. *Int J Radiat Biol* 2003; **79**: 129–36.
37. Moore RJ, Owens DM, Stamp G, Arnott C, Burke F, East N, Holdsworth H, Turner L, Rollins B, Pasparakis M, Kollias G, Balkwill F. Mice deficient in tumor necrosis factor-alpha are resistant to skin carcinogenesis. *Nat Med* 1999; **5**: 828–31.
38. Suganuma M, Okabe S, Marino MW, Sakai A, Sueoka E, Fujiki H. Essential role of tumor necrosis factor alpha (TNF-alpha) in tumor promotion as revealed by TNF-alpha-deficient mice. *Cancer Res* 1999; **59**: 4516–8.
39. Voronov E, Shouval DS, Krelin Y, Cagnano E, Benharroch D, Iwakura Y, Dinarello CA, Apte RN. IL-1 is required for tumor invasiveness and angiogenesis. *Proc Natl Acad Sci USA* 2003; **100**: 2645–50.
40. Bernert H, Sekikawa K, Radcliffe RA, Iraqi F, You M, Malkinson AM. Tnfa and Il-10 deficiencies have contrasting effects on lung tumor susceptibility: gender-dependent modulation of IL-10 haploinsufficiency. *Mol Carcinog* 2003; **38**: 117–23.
41. Suganuma M, Okabe S, Kurusu M, Iida N, Ohshima S, Saeki Y, Kishimoto T, Fujiki H. Discrete roles of cytokines, TNF-alpha, IL-1, IL-6 in tumor promotion and cell transformation. *Int J Oncol* 2002; **20**: 131–6.
42. Hudson JD, Shoaibi MA, Maestro R, Carnero A, Hannon GJ, Beach DH. A proinflammatory cytokine inhibits p53 tumor suppressor activity. *J Exp Med* 1999; **190**: 1375–82.
43. Hamerman JA, Ogasawara K, Lanier LL. Cutting edge: toll-like receptor signaling in macrophages induces ligands for the NKG2D receptor. *J Immunol* 2004; **172**: 2001–5.
44. Nakamoto Y, Guidotti LG, Kuhlen CV, Fowler P, Chisari FV. Immune pathogenesis of hepatocellular carcinoma. *J Exp Med* 1998; **188**: 341–50.
45. El-Omar EM, Carrington M, Chow WH, McColl KE, Bream JH, Young HA, Herrera J, Lissowska J, Yuan CC, Rothman N, Lanyon G, Martin M, Fraumeni JF Jr, Rabkin CS. Interleukin-1 polymorphisms associated with increased risk of gastric cancer. *Nature* 2000; **404**: 398–402.
46. Neriishi K, Nakashima E, Delongchamp RR. Persistent subclinical inflammation among A-bomb survivors. *Int J Radiat Biol* 2001; **77**: 475–82.
47. Kusunoki Y, Kyoizumi S, Hirai Y, Suzuki T, Nakashima E, Kodama K, Seyama T. Flow cytometry measurements of subsets of T, B and NK cells in peripheral blood lymphocytes of atomic bomb survivors. *Radiat Res* 1998; **150**: 227–36.
48. Kidd P. Th1/Th2 balance: the hypothesis, its limitations, and implications for health and disease. *Altern Med Rev* 2003; **8**: 223–46.
49. Mathew A, Ennis FA, Rothman AL. Transient decreases in human T cell proliferative responses following vaccinia immunization. *Clin Immunol* 2000; **96**: 100–7.
50. Papadaki HA, Coulocheri S, Eliopoulos GD. Patients with chronic idiopathic neutropenia of adults have increased serum concentrations of inflammatory cytokines and chemokines. *Am J Hematol* 2000; **65**: 271–7.
51. Bursill CA, Cai S, Channon KM, Greaves DR. Adenoviral-mediated delivery of a viral chemokine binding protein blocks CC-chemokine activity *in vitro* and *in vivo*. *Immunobiology* 2003; **207**: 187–96.
52. Jinquan T, Jing C, Jacobi HH, Reimert CM, Millner A, Quan S, Hansen JB, Dissing S, Malling HJ, Skov PS, Poulsen LK. CXCR3 expression and activation of eosinophils: role of IFN-gamma-inducible protein-10 and monokine induced by IFN-gamma. *J Immunol* 2000; **165**: 1548–56.
53. de Lemos JA, Morrow DA, Sabatine MS, Murphy SA, Gibson CM, Antman EM, McCabe CH, Cannon CP, Braunwald E. Association between plasma levels of monocyte chemoattractant protein-1 and long-term clinical outcomes in patients with acute coronary syndromes. *Circulation* 2003; **107**: 690–5.
54. Cornelli U, Terranova R, Luca S, Cornelli M, Alberti A. Bioavailability and antioxidant activity of some food supplements in men and women using the D-Roms test as a marker of oxidative stress. *J Nutr* 2001; **131**: 3208–11.
55. Wallstrom P, Frenkel K, Wirfalt E, Gullberg B, Karkoszka J, Seidegard J, Janzon L, Berglund G. Antibodies against 5-hydroxymethyl-2′-deoxyuridine are associated with lifestyle factors and GSTM1 genotype: a report from the Malmo Diet and Cancer cohort. *Cancer Epidemiol Biomarkers Prev* 2003; **12**: 444–51.
56. Kondo S, Toyokuni S, Tanaka T, Hiai H, Onodera H, Kasai H, Imamura M. Overexpression of the hOGG1 gene and high 8-hydroxy-2′-deoxyguanosine (8-OHdG) lyase activity in human colorectal carcinoma: regulation mechanism of the 8-OHdG level in DNA. *Clin Cancer Res* 2000; **6**: 1394–400.
57. Hagenlocher T, Nair J, Becker N, Korfmann A, Bartsch H. Influence of dietary fatty acid, vegetable, and vitamin intake on etheno-DNA adducts in white blood cells of healthy female volunteers: a pilot study. *Cancer Epidemiol Biomarkers Prev* 2001; **10**: 1187–91.

IJE vol.33 no.3 © International Epidemiological Association 2004; all rights reserved.
Advance Access publication 22 April 2004

International Journal of Epidemiology 2004;33:485–492
DOI: 10.1093/ije/dyh097

Improving the efficiency of nested case-control studies of interaction by selecting controls using counter matching on exposure

John B Cologne,[1] Gerald B Sharp,[2] Kazuo Neriishi,[3] Pia K Verkasalo,[5] Charles E Land[6] and Kei Nakachi[4]

Accepted 17 December 2003

Background Studies of the effect of exposure to a risk factor measured in an entire cohort may be augmented by nested case-control subsets to investigate confounding or effect modification by additional factors not practically assessed on all cohort members. We compared three control-selection strategies—matching on exposure, counter matching on exposure, and random sampling—to determine which was most efficient in a situation where exposure is a known, continuous variable and high doses are rare.

Methods We estimated the power to detect interaction using four control-to-case ratios (1:1, 2:1, 4:1, and 8:1) in a planned case-control study of the joint effect of atomic bomb radiation exposure and serum oestradiol levels on breast cancer. Radiation dose is measured in the entire cohort, but because neither serum oestradiol level nor the true degree of interaction was known, we simulated values of oestradiol and hypothetical levels of oestradiol–radiation interaction.

Results Compared with random sampling, power to detect interaction was similarly higher with either matching or counter matching with two or more controls.

Conclusions Because counter matching is generally at least as efficient as random sampling, whereas matching on exposure can result in loss of efficiency and precludes estimation of exposure risk, we recommend counter matching for selecting controls in nested case-control studies of the joint effects of multiple risk factors when one is previously measured in the full cohort.

Keywords Nested case-control studies, probability sample, matching, counter matching, breast cancer, radiation effects

In cohort studies aimed at investigating the effects of already measured risk factors—such as radiation exposure in the survivors of the atomic bombings of Hiroshima and Nagasaki, Japan—nested case-control studies may be conducted to analyse the effects of additional factors that cannot be assessed practically in the entire cohort. The purpose might be to study the effects of potential confounders or effect-modifying factors on the exposure risk. If the exposure is rare or has a skewed distribution, ignoring it in selecting controls can lead to a loss of statistical efficiency, so exposure-based methods of control selection might be considered.

As an example, although radiation and oestradiol are both strong risk factors for breast cancer, only radiation dose is known for atomic bomb survivors. The relative risk of early-onset breast cancer (diagnosis under age 35) for 1 Sv of radiation is 14 among women irradiated by the atomic bombs before age 20; the overall relative risk of breast cancer ranges from 2.3 to 3.4 among all women exposed under the age of 40.[1] Key, Verkasalo, and Banks showed that serum oestradiol is

Departments of [1]Statistics, [2]Epidemiology, [3]Clinical Studies, and [4]Radiobiology/Molecular Epidemiology, Radiation Effects Research Foundation, Hiroshima, Japan.

[5] Unit of Environmental Epidemiology, National Public Health Institute, Finland.

[6] Radiation Epidemiology Branch, National Cancer Institute, US.

Correspondence: John B Cologne, Department of Statistics, Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami-ku, Hiroshima 732-0815, Japan. E-mail: cologne@rerf.or.jp

positively associated with risk of breast cancer, with risks being about twice as high for postmenopausal women with high, as opposed to low, serum oestradiol concentrations.[2] Little is known, however, about oestradiol levels and the risk of pre-menopausal breast cancer. Furthermore, the joint effect of radiation and oestradiol has not been studied, although Land *et al.*[3] demonstrated interactions between radiation and breast cancer risk factors that may be related to constitutional hormone levels. At the Radiation Effects Research Foundation (RERF), we are conducting a study of radiation and oestradiol as joint risk factors for pre-menopausal breast cancer using stored sera obtained from atomic bomb survivors who participated in biennial clinical examinations conducted for RERF's Adult Health Study. Oestradiol, which is expensive to assay and requires sera, for which supplies are limited, will be measured in all cases but only a subset of controls. This raises the issue of how best to select controls to provide maximum statistical efficiency.

Selecting controls by individually matching them to cases on radiation exposure can improve statistical efficiency for testing interaction with another factor.[4] If there is evidence of interaction or possible confounding in the case-control sample, a logical next step in the analysis would be to examine how exposure risk estimates vary with the level of the other factor. Matching on exposure allows studying the effect of confounder/effect-modifier *per se* but precludes studying its effect on the exposure risk without additional information, such as comes from the cohort.[3] An alternative to matching is weighted sampling of controls using counter-matching, where controls are selected to fill exposure strata not occupied by the case.[5–8] Counter matching also allows estimation of the exposure risk, and the efficiency for studying both confounding and effect modification can be improved relative to random sampling of controls.[9] Furthermore, counter matching allows the investigator to fix the number of controls in advance and is easily implemented with prospective, risk-set based selection. Many of the references on these designs provide justification and intuitive explanation as to why exposure-based sampling is efficient.

Counter matching has been shown to generate better efficiency for testing interaction than matching over a wide range of exposure risks and degrees of correlation between exposure and another risk factor when both are dichotomous,[8] but comparisons have not been made for continuous risk factors, such as radiation dose and oestradiol. The proposed breast cancer study provides a basis for making that comparison in the case of a rare exposure that may interact positively with an additional factor, the situation in which matching achieves the greatest gain in efficiency.[4] Our objective was to assess the extent to which matching and counter matching impact statistical power for detecting interaction relative to random control selection.

Subjects and Methods

Study design

We are conducting a nested case-control study of pre-menopausal breast cancer, radiation exposure, and serum oestradiol levels using all currently available cases.[10] Radiation doses are known, but oestradiol is to be measured in the case-control subset using stored sera. Day of menstrual cycle at the time of serum collection is not known; therefore it was decided to use the average of two measurements on serum specimens collected on different occasions. There were 80 cases of pre-menopausal breast cancer and 5644 cancer-free women with stored serum. Radiation doses were calculated according to Dosimetry System 1986 (DS86).[11] Radiation doses to the breast were calculated in Sievert (Sv), combining gamma and neutron components with a relative weight of 10 for neutrons.

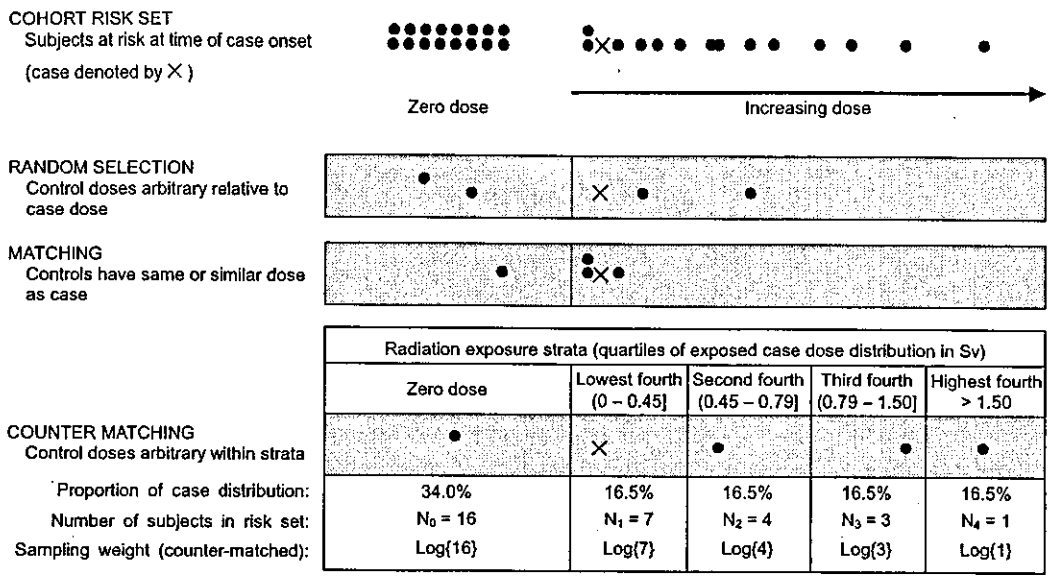
To assess the power of the study for detecting interaction, we simulated oestradiol levels and their interaction with radiation. We then calculated the resulting power using one, two, four, or eight controls selected from case risk-sets using three approaches: (1) random sampling, (2) matching as closely as possible on radiation exposure, or (3) counter matching on radiation exposure.

Selecting controls by random sampling is simple. Within risk sets defined by age, date, and availability of stored serum, controls are selected at random, without regard for exposure status. Matching on radiation exposure is also straightforward; within each risk set, the potential controls whose exposure values are closest to the case's exposure are selected. Note, however, that matching on exposure in addition to matching on other factors can be more complicated.[12] If there are more tied values among the potential controls than the number needed, the controls are selected at random from among the tied subjects.

With counter matching, exposure strata are defined based on the number of controls to be selected per case and on the distribution of exposure among the cases. In each risk set one control is selected from each of the exposure strata not occupied by the case. In the study described here, there were many people with dose zero; we therefore defined the lowest exposure category to be zero. The other categories were determined by the appropriate percentiles of the distribution of exposure values among the exposed cases. For example, with eight controls per case there were nine exposure strata: zero plus eighths defined by octiles of the case exposure values (cutpoints: 0.14, 0.45, 0.66, 0.79, 1.15, 1.50, and 2.04 Sv; Supplementary Material: Distribution of Radiation Doses). With four controls per case there were five strata: zero plus fourths defined by quartiles of the case exposures (cutpoints: 0.45, 0.79, and 1.50 Sv; Figure 1).

Simulation of serum oestradiol levels and interaction with radiation

Oestradiol values were computer generated to have overall mean and variance equal to those from a previous study performed in the same cohort.[13] Because there was no information about day of menstrual cycle on which serum was collected, we accounted for day-to-day variation by mimicking random sampling from the menstrual cycle using data for British women from Verkasalo and her associates.[14] First, a random integer representing day of cycle was obtained using a uniform random number generator. Then, a random $\log_{10}$-oestradiol value was obtained by generating a random Normal variable with mean and variance equal to those for the same cycle day among the British study group. Finally, the resulting values were adjusted to have overall (not day-specific) mean and variance equal to those found in the previous atomic bomb



Radiation exposure strata (quartiles of exposed case dose distribution in Sv)				
Zero dose	Lowest fourth (0 – 0.45]	Second fourth (0.45 – 0.79]	Third fourth (0.79 – 1.50]	Highest fourth > 1.50
Proportion of case distribution:	34.0%	16.5%	16.5%	16.5%	16.5%
Number of subjects in risk set:	$N_0 = 16$	$N_1 = 7$	$N_2 = 4$	$N_3 = 3$	$N_4 = 1$
Sampling weight (counter-matched):	Log{16}	Log{7}	Log{4}	Log{3}	Log{1}

Figure 1 Illustration of control-selection strategies with four controls per case. The case exposure value (×) in this illustration is close to zero and in the lowest of the fourths derived from quartiles among non-zero-dose cases. With random sampling, controls would be selected randomly without regard to their dose. With matching, the four controls with doses closest to that of the case would be selected. With counter matching, one control would be selected randomly from each of the non-case strata (the remaining three exposure fourths and the zero-dose stratum)

survivor study. This adjustment was made separately for cases and controls. The simulated values were standardized to produce a log odds ratio of $\varphi_A = 0.359$, the result obtained with an average $\log_{10}$-oestradiol difference of 7% between cases and controls in the previous study, where controls were matched to cases on radiation dose. This corresponds to odds of disease of 1.43 for a 1.41-fold higher oestradiol level. For each subject we used the average of two simulated oestradiol values from arbitrary cycle days to reduce the error due to lack of knowledge of cycle day on which the serum was collected.

Because radiation dose and case status were already known in the cohort, we simulated the interaction by adjusting the mean of the Normal distribution from which case $\log_{10}$-oestradiol values were generated so that the average case-control mean difference in $\log_{10}$-oestradiol level increased linearly with radiation dose, while $\log_{10}$-oestradiol values among the controls were generated independently of radiation dose. The dose-dependent case $\log_{10}$-oestradiol means were calculated to produce log odds ratios (OR) according to the following model:

$$\varphi(d) = \varphi_A \times f(d) = \varphi_A \times \left[1 + \gamma\left(\frac{d - 0.75}{0.75}\right)\right],$$

so that the log OR is φ_A at a dose of 0.75 Sv, close to the average dose among the cases. Three levels of interaction were selected to produce low, moderate, and high power of detection in the full cohort. These were simulated by setting γ so that, with a doubling of dose to 1.5 Sv, the relative changes in the log OR for $\log_{10}$-oestradiol with interaction were 0.5 (low degree of interaction), 0.75 (moderate degree of interaction), or 1.0 (high degree of interaction); Figure 2. The low degree of interaction results in a log OR of 0.5φ (OR = 1.20) at 0 Sv and a log OR of 1.5φ (OR = 1.71) at twice the median dose (1.5 Sv). The moderate degree of interaction results in corresponding log OR of 0.25φ and 1.75φ (OR = 1.09 and 1.87). The high degree

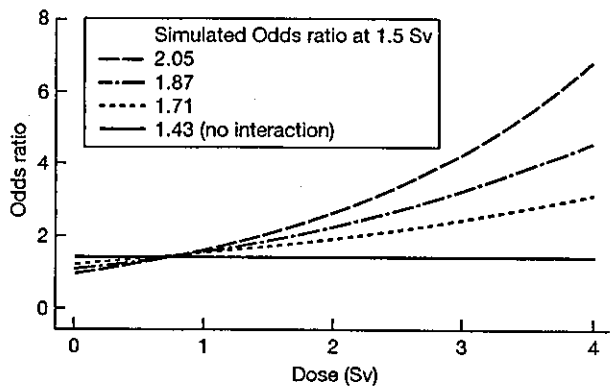


Figure 2 The three levels of hypothetical oestradiol-radiation interaction used in the simulations. Plots show the odds ratio for oestradiol as a function of radiation dose. The overall odds ratio (1.43) for the case-control difference of 7% in log oestradiol value was set to occur at a dose of 0.75 Sv, near the median dose of the cases, corresponding to the result of the previous study

of interaction results in corresponding log OR of 0 and 2.0φ (OR = 1.00 and 2.05); it represents extreme effect modification in that it results in no oestradiol effect (OR = 1.0) among women not exposed to radiation.

We simulated 500 random case-control study outcomes for each of the 12 configurations defined by the number of controls per case selected (1, 2, 4, or 8) and the degree of interaction assumed (low, moderate, or high). All simulated data were generated using S-plus (MathSoft Inc., Seattle, Washington).

Statistical analysis of simulated data

We analysed the counter-matched design according to Langholz and Borgan,[5] using conditional logistic regression with sampling weights as offsets.[15] (An offset is added to the logistic

regression model by entering it as a covariate with coefficient fixed at 1.) Counter-matched sampling weights were calculated separately for each risk set (i, $i = 1, \ldots, I$), and exposure stratum (j, $j = 1, \ldots, J$) as:

$$\omega_{ij} = \log\{\text{number of cohort subjects in risk set } i \text{ and exposure stratum } j\}.$$

Weights are calculated in the same way for both cases and controls (Figure 1).

We studied effect modification via statistical interaction by fitting the model:

$$f(d, e) = \beta d + \theta e + \gamma de$$

where $f(d, e)$ is the log odds of breast cancer, β is the log OR for a unit (1 Sv) difference in radiation dose (d), θ is the log OR for a unit difference in $\log_{10}$-oestradiol (e), and γ is the interaction parameter. Interaction was tested by the χ^2 approximation to the likelihood ratio test of the null hypothesis $\gamma = 0$, which was rejected if the test statistic exceeded 3.84. Note that this is equivalent to failure of the likelihood-based CI to include the null value $\gamma = 0$. We calculated power as the proportion of 500 simulations in which the null hypothesis of no interaction was rejected. Because the power of a cohort based analysis—if oestradiol levels were known for all cohort members—would increase with increasing degree of interaction, we compared the power of each case-control design to that of the entire cohort. For each level of interaction we simulated 1000 cohort oestradiol values.

Counter-matched case-control samples were analysed using conditional logistic regression with weights as described above. Matched and randomly selected case-control samples were analysed using conditional logistic regression without weights. Cohort data were analysed using unconditional logistic regression. All analyses were conducted using Epicure (Hirosoft Inc., Seattle, Washington).

Results

Figure 3 shows the distribution of randomly generated pre-menopausal oestradiol values for 80 cases and 160 controls from one set of 500 simulations assuming no interaction with radiation. There was substantial variation in simulated oestradiol levels within day and due to day of cycle. Oestradiol levels were 16% higher on average ($\log_{10}$-oestradiol values were 7% higher) in cases than in controls.

As shown in Table 1, compared with random selection of controls, the counter-matched and matched strategies increased power to detect interaction for each level of interaction when at least two controls per case were selected. All three methods demonstrated similar levels of power relative to the full cohort (slightly less than 40%) with only one control per case. Relative differences in power between the three strategies were similar regardless of the level of interaction. In Figure 4 we summarize the power of each design using curves fit to the average power (averaged over levels of interaction). Random control selection resulted in power that ranged from 50% of the cohort level with two controls per case to slightly more than 80% with eight controls per case. On the other hand, the exposure-based sampling strategies achieved levels of power ranging from almost 80% with two controls per case to 100% (maximum) power with eight controls per case. For random sampling to achieve an acceptable level of power equivalent to that of the exposure-based control-selection designs required about twice as many controls, indicating that the exposure-based designs were about twice as efficient as random sampling for detecting interaction.

Repeated runs with counter matching and four controls per case with the moderate degree of interaction resulted in absolute power estimates of 0.466, 0.516, 0.518, 0.474, and 0.530 (average: 0.501, standard deviation 0.029). Because the variance of a binomial variable is a maximum when the proportion is 0.5, the other simulation results would be equally, or more, precise. The differences between exposure-related sampling designs and random sampling were much larger than this repeated simulation variability. However, differences between power for counter matching and matching were small and on the order of the simulation precision.

Conclusions and Discussion

Case-control studies of interaction between multiple risk factors generally have low statistical power,[16] so designs that are suitably sized for studying main effects may be inadequate for studying effect modification.[17] The two exposure-related control-selection strategies studied here (matching and counter matching) resulted in similar gains in efficiency for testing interaction when two or more controls were selected per case. We conclude that, while more than 10 controls per case would be needed to achieve greater than 90% of the maximum, cohort level of power with random control selection, only 5–6 controls—about half as many as with random sampling—would be needed using either of the exposure-based control-selection designs. In practice, there could be missing specimens or refusal of informed consent, so using more than 5–6 controls per case might be considered based on the specific application. We do not recommend the use of only one control per case for studies of interaction.

The risk factors in our investigation were assumed to have positive interaction and one was a rare exposure already measured in the full cohort. This is the situation where matching performs best; in more general situations—i.e., when the two risk factors do not interact positively or when the matching factor is not rare—matching can lead to a loss of efficiency.[4] Counter matching generally improves statistical efficiency for studying interaction[9] and, unlike matching, further allows studying the exposure risk with adjustment for the other factor measured only in the case-control sample. Counter matching using sampling within risk-set strata is no more difficult to perform than matching on exposure, which can be complicated when additional risk-set matching factors are involved,[11] and both strategies require the use of conditional logistic regression. Counter matching additionally requires sampling weights, which are calculated from the numbers of cohort subjects in each risk-set exposure stratum in the cohort. Being able to examine the adjusted exposure risk would usually outweigh the extra effort involved in calculating the weights. We conclude that counter matching, and not matching, should generally be used to increase efficiency if a nested case-control study of joint effects is planned when one risk factor is known in the cohort.

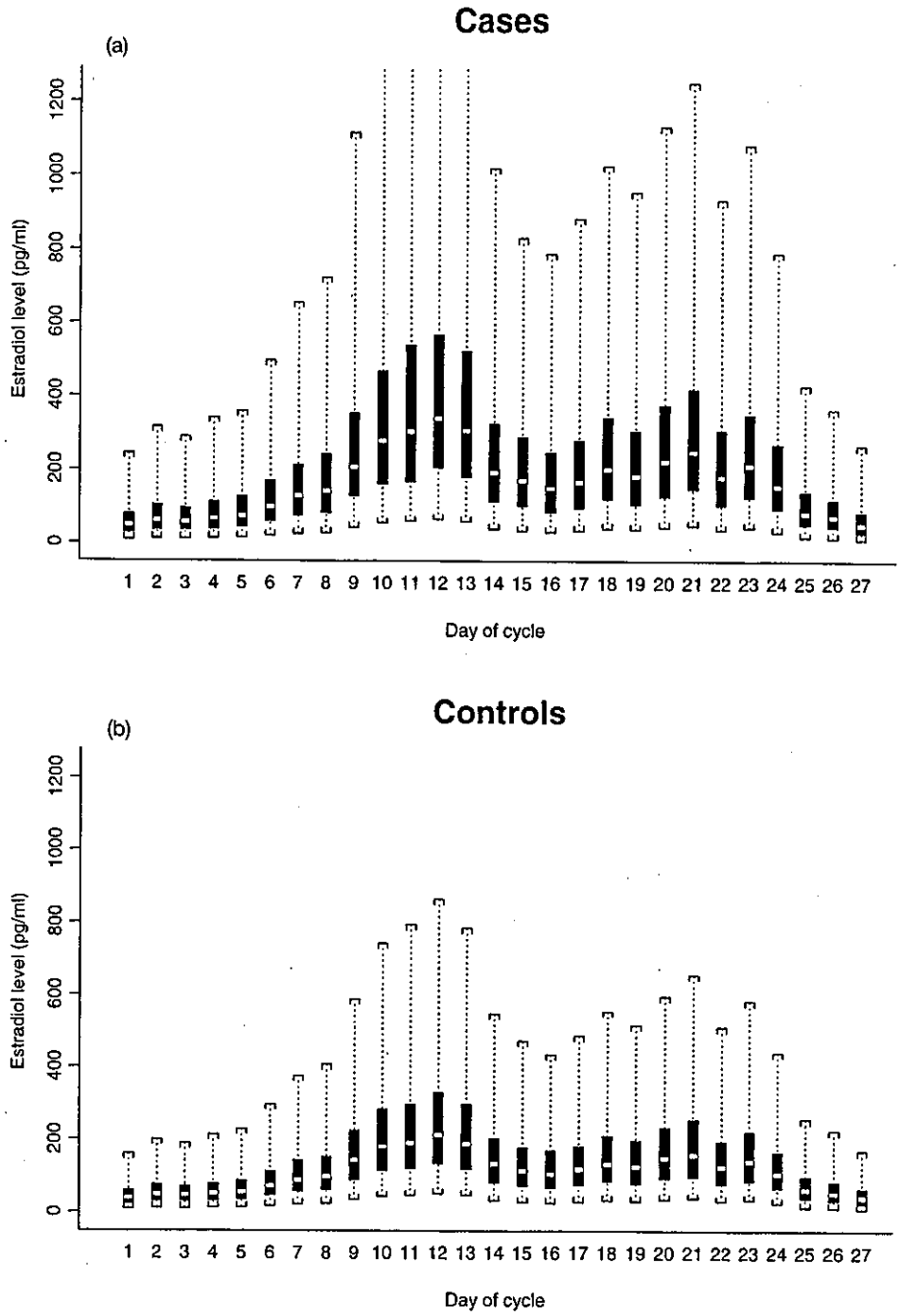


Figure 3 Boxplots illustrating computer-generated oestradiol values by day of cycle among pre-menopausal breast cancer (a) cases and (b) selected controls. Data are from 500 simulations with 80 cases and 2 controls per case. Day of cycle is defined with 0 being 28 days prior to the start of the subsequent cycle (i.e. day 28 is the start of the next cycle). Case values are 16% higher on average than control values. The boxplots show the median (white dot), inter-quartile range (filled bar), and the most extreme observations (brackets)

When calculating the power of a study involving sampling from a cohort, two issues deserve consideration: power of the full cohort and power of the study design. There is little point in selecting a subset to investigate interaction if even the cohort is too small or the effect too weak to provide sufficient statistical power. If the cohort has sufficient power, then the question becomes what type and size of design will provide the greatest possible efficiency within the limitations of financial cost, time, biological specimen availability, and other considerations. Breslow and Day point out that some sampling of cohort risk

Table 1 Estimated power for detecting interaction relative to the full cohort for three control-selection strategies

Degree of interaction (cohort power)	Control-selection strategy	Control-to-case ratio 1	2	4	8
Low (0.293)	Random	0.40	0.54	0.75	0.92
	Matched	0.37	0.80	1.00	1.00
	Counter matched	0.36	0.78	1.00	1.00
Moderate (0.548)	Random	0.32	0.47	0.73	0.82
	Matched	0.34	0.77	1.00	1.00
	Counter matched	0.38	0.80	0.85	1.00
High (0.778)	Random	0.37	0.51	0.69	0.86
	Matched	0.35	0.70	0.96	1.00
	Counter matched	0.38	0.77	0.89	1.00

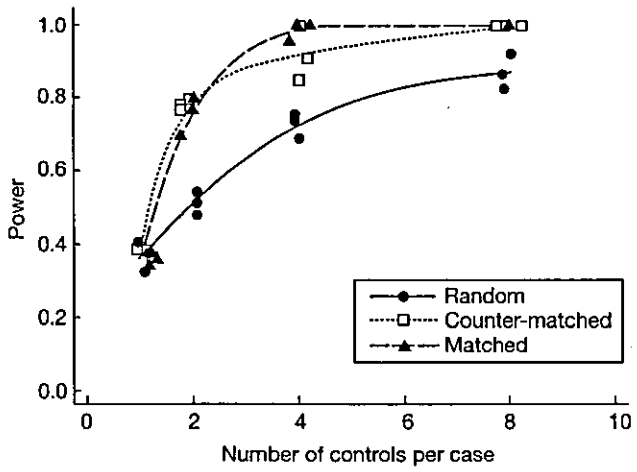


Figure 4 Power of the control-sampling strategies for detecting interaction relative to the power of the full cohort. Points are values from Table 1, jittered slightly on the abscissa to reduce overlap. Lines are the results of fitting cubic power curves for purposes of smoothing the means (averaged over level of interaction)

sets can generally be performed with little loss of efficiency.[18] We have not considered the trade-off between cost and benefit here (see, for example, Reilly[19]), but in designing studies investigators must decide how much of the cohort power they are willing to sacrifice to achieve the necessary logistical savings. We have not investigated all possible designs, but for two approaches to nested case-control selection with fixed risk-set size, we have demonstrated that using counter matching can allow the researcher to achieve the same level of efficiency using about half as many controls as would be needed if controls were selected randomly.

The nested case-control design allows repeated selection of subjects in different risk sets; even cases can serve as controls in risk sets prior to their disease onset. Thus, there can be greater efficiency (in terms of number of subjects) depending on how many subjects are selected repeatedly by chance. In our application, the number of potential controls was large compared with the number of risk sets, so the probability of repeated selection was small. The total number of subjects needed for a study will depend on this ratio as well as on the random draw of subjects. With matching, the number of potential controls at rare levels of exposure is limited and may lead to repeated selection. However, when counter matching is used, because dose strata are defined by quantiles, repeated selection is not likely to occur except for very large control:case ratios.

In studies with exposure known in the entire cohort, there is additional information on exposure risk in the non-selected subjects. Two-stage designs can improve efficiency.[20–22] Langholz and Goldstein[23] proposed a likelihood for analysing the case-control data only using a proportional odds model with multi-stage sampling. Land and others[3] proposed a method for incorporating the cohort risk estimate into the analysis of case-control subsets matched on exposure using more general risk models. There are also alternatives to the nested case-control design with counter matching. Borgan *et al.* addressed exposure-stratified selection in the case-cohort design.[24] Randomized recruitment as an alternative to counter matching can also result in efficiency gains.[25] Much remains to be done to synthesize the various designs and methods of analysis, but that is beyond the scope of the present work.

Because the present investigation was based on our interest in effect modification, we had to speculate as to what form it might take in order to simulate study power. Huang *et al.* reported that risk of breast cancer for medical irradiation to the chest in pre-/peri-menopausal women tended to be associated with oestrogen-receptor negative tumours,[26] suggesting that mechanisms other than those dependent on hormonal exposure may be involved. Radiation might cause additional genetic alterations that result in more rapid progression of breast cancer associated with oestrogen receptor negative phenotype. Oestrogen receptor negative breast cancer cells have been reported to be relatively resistant to IL-6 induced apoptosis,[27] so they may be more proliferative. If radiation induced alterations in signal transduction systems that were independent of the oestrogen receptor signalling system, then the joint effect of radiation and oestradiol could be multiplicative. If such alterations were dependent on the oestrogen receptor signalling system, then the joint effect could be greater than multiplicative. On the other hand, radiation exposure may lead to early onset of menopause,[28] which could indirectly reduce the risk of breast cancer by decreasing the duration of exposure to constitutional estrogens. Therefore,

in studying the joint effects of radiation and oestradiol, it is important to have sufficient power to detect or rule out interaction on the multiplicative scale to facilitate the planning of in-depth mechanistic studies.

Because these possibilities for interaction are mostly speculative, there was, in the present study, no basis to assume any particular type of effect modification between radiation and oestradiol. We therefore studied several arbitrary degrees of statistical interaction using a log-linear model. Effect modification can take other forms, including interaction on an additive scale. Such statistical interactions have been defined as effect-measure modification as distinguished from true effect modification, or biological interaction,[29,30] which implies that the joint effect of multiple risk factors exceeds the sum of their individual risks. In the analysis of data from a nested case-control study, one should consider alternatives to the standard log-linear logistic-regression model for the joint effect of multiple risk factors, such as additive or mixture models.[31]

In summary, we have demonstrated that matching and counter-matching on a known, continuous exposure variable provide equal gains in statistical power in a nested case-control study of risk-factor interaction with a control:case ratio of at least 2:1. However, matching on exposure prevents studying the effect of exposure after adjusting for one or more other risk factors which might confound or modify the exposure risk, study aspects that counter matching addresses with greater efficiency than random sampling. We conclude that counter matching is superior to both matching on exposure and random control selection for nested case-control studies of effect modification when there is a known exposure.

Acknowledgements

Special thanks go to Professor Bryan Langholz for advice and instruction regarding the method of counter matching. The study would not have been possible without the skilful technical support of Ms Sachiyo Funamoto. This publication was supported in part by research protocol 6–02 of the Radiation Effects Research Foundation (RERF), Hiroshima and Nagasaki, Japan. RERF is a private, non-profit foundation funded by the Japanese Ministry of Health, Labour and Welfare and the US Department of Energy, the latter through the National Academy of Sciences. The authors also acknowledge the support of Grants-in-Aid Nos. 14580356 and 14031227 from the Ministry of Education, Culture, Sports, Science, and Technology of Japan and grant number NCI-4893–8–001 from the US National Institutes of Health.

KEY MESSAGES

- Selecting controls with consideration of a known, rare exposure can improve efficiency over random control selection in nested case-control studies of interaction.
- Counter matching produces the same gain in efficiency as matching for such studies but affords greater flexibility.
- Implementation of counter matching and estimation of power of a nested case-control study are illustrated using a case study of breast cancer and two risk factors, one of which is known in the cohort.

References

[1] Tokunaga M, Land CE, Tokuoka S, Nishimori I, Soda M, Akiba S. Incidence of female breast cancer among atomic bomb survivors, 1950–1985. *Radiat Res* 1994;**138**:209–23.

[2] Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. *Lancet Oncol* 2001;**2**:133–40.

[3] Land CE, Hayakawa N, Machado SG *et al.* A case-control interview study of breast cancer among Japanese A-bomb survivors. II. Interactions with radiation dose. *Cancer Causes Control* 1994;**5**:167–76.

[4] Thomas DC, Greenland S. The efficiency of matching in case-control studies of risk-factor interactions. *J Chron Dis* 1985;**38**:569–74.

[5] Langholz B, Borgan Ø. Counter-matching: A stratified nested case-control sampling method. *Biometrika* 1995;**82**:69–79.

[6] Cologne JB. Counterintuitive matching (editorial). *Epidemiology* 1997;**8**:227–29.

[7] Steenland K, Deddens JA. Increased precision using countermatching in nested case-control studies. *Epidemiology* 1997;**8**:238–42.

[8] Cologne J, Langholz B. Selecting controls for assessing interaction in nested case-control studies. *J Epidemiol* 2003;**13**:184–93.

[9] Langholz B, Goldstein L. Risk set sampling in epidemiologic cohort studies. *Statist Sci* 1996;**11**:35–53.

[10] Sharp GB, Neriishi K, Hakoda M *et al.* A nested case-control study of breast and endometrial cancer in the cohort of Japanese atomic bomb survivors. Research Protocol 6-02: Radiation Effects Research Foundation, Hiroshima, Japan; 2002.

[11] Roesch WC (ed.). *US-Japan Joint Reassessment of Atomic Bomb Radiation Dosimetry in Hiroshima and Nagasaki.* Hiroshima, Japan: Radiation Effects Research Foundation, 1975.

[12] Cologne JB, Shibata Y. Optimal case-control matching in practice. *Epidemiology* 1995;**6**:221–25.

[13] Kabuto M, Akiba S, Stevens RG, Neriishi K, Land CE. A prospective study of estradiol and breast cancer in Japanese women. *Cancer Epidemiol Biomarkers Prev* 2000;**9**:575–79.

[14] Verkasalo PK, Thomas HV, Appleby PN, Davey GK, Key TJ. Circulating levels of endogenous hormones and their relations to risk factors for breast cancer: a cross-sectional study in 1092 pre- and postmenopausal women (United Kingdom). *Cancer Causes Control* 2001;**12**:47–59.

[15] McCullagh P, Nelder JA. *Generalized Linear Models. 2nd Edn.* London: Chapman and Hall, 1989.

[16] Greenland S. Tests for interaction in epidemiologic studies: a review and a study of power. *Stat Med* 1983;**2**:243–51.

[17] Smith PG, Day NE. The design of case-control studies: The influence of confounding and interaction effects. *Int J Epidemiol* 1984;**13**:356–65.

[18] Breslow NE, Day NE. *Statistical Methods in Cancer Research. Volume II—The Design and Analysis of Cohort Studies*. Lyon: International Agency for Research on Cancer, 1987, pp. 200–02.

[19] Reilly M. Optimal sampling strategies for two-stage studies. *Am J Epidemiol* 1996;**143**:92–100.

[20] Breslow NE, Cain KC. Logistic regression for two stage case-control data. *Biometrika* 1988;**75**:11–20.

[21] Zhao LP, Lipsitz S. Designs and analysis of two-stage studies. *Stat Med* 1992;**11**:769–82.

[22] Breslow NE, Chatterjee N. Design and analysis of two-phase studies with binary outcome applied to Wilms tumour prognosis. *Appl Statist* 1999;**48**:457–68.

[23] Langholz B, Goldstein L. Conditional logistic analysis of case-control studies with complex sampling. *Biostatistics* 2001;**2**:63–84.

[24] Borgan Ø, Langholz B, Samuelsen SO, Goldstein L, Pogoda J. Exposure stratified case-cohort designs. *Lifetime Data Analysis* 2000;**6**:39–58.

[25] Weinberg CR, Wacholder S. The design and analysis of case-control studies with biased sampling. *Biometrics* 1990;**46**:963–75.

[26] Huang W-Y, Newman B, Millikan RC, Schnell MJ, Hulka BS, Moorman PG. Hormone-related factors and risk of breast cancer in relation to estrogen receptor and progesterone receptor status. *Am J Epidemiol* 2000;**151**:703–14.

[27] Chiu JJ, Sgagias MK, Cowan KH. Interlukin 6 acts as a paracrine growth factor in human mammary carcinoma cell lines. *Clin Cancer Res* 1996;**2**:215–21.

[28] Soda M, Cologne J. Radiation-accelerated age at menopause. *RERF Update* 1993;**5**:5–6.

[29] Rothman KJ. *Epidemiology: An Introduction*. Oxford: Oxford University Press, 2002; p. 170.

[30] Greenland S, Rothman KJ. Concepts of interaction. In: Rothman KJ, Greenland S (eds). *Modern Epidemiology*. Philadelphia: Lippincott Williams & Wilkins, 1998; pp. 329–42.

[31] Thomas DC. General relative-risk models for survival time and matched case-control analysis. *Biometrics* 1981;**37**:673–86.

Journal of Pathology
J Pathol 2004; **203**: 545–550
Published online 18 February 2004 in Wiley InterScience (www.interscience.wiley.com). **DOI**: 10.1002/path.1542

Original Paper

Allelic length of a CA dinucleotide repeat in the *egfr* gene correlates with the frequency of amplifications of this sequence — first results of an inter-ethnic breast cancer study

Horst Buerger,[1]* Jens Packeisen,[2] Almuth Boecker,[3] Nicola Tidow,[3] Christian Kersting,[1] Krzysztof Bielawski,[4] Jorma Isola,[5] Yasushi Yatabe,[6] Kei Nakachi,[7] Werner Boecker[1] and Burkhard Brandt[3]

[1]*Institute of Pathology, University of Muenster, Germany*
[2]*Institute of Pathology, Osnabrück, Germany*
[3]*Institute of Clinical Chemistry and Laboratory Medicine, University of Muenster, Germany*
[4]*Department of Biotechnology, University of Gdansk, Poland*
[5]*Institute of Medical Technology, University of Tampere, Finland*
[6]*Aichi Cancer Center Research Institute, Nagoya, Japan*
[7]*Department of Epidemiology, Saitama Cancer Center, Japan*

*Correspondence to:
Horst Buerger, MD, PhD,
Institute of Pathology,
Westfälische Wilhelmsuniversität Münster,
Albert-Schweitzer-Strasse 33,
48149 Münster, Germany.
E-mail: burgerh@uni-muenster.de

Received: 29 January 2003
Revised: 26 November 2003
Accepted: 8 December 2003

Abstract

Overexpression of the epidermal growth factor receptor (EGFR) is a common finding in invasive breast cancer and represents a potential target for new treatment options. However, little is known about the parameters that might indicate a potential clinical response for these anti-EGFR-based therapies. In order to gain further insights into the interplay between the length of a CA-SSR I repeat in intron 1 of *egfr*, copy numbers of this untranslated regulatory sequence, and protein expression, the present study investigated breast cancers from Germans and Japanese patients by microsatellite analysis, quantitative 5′ nuclease assay by *egfr* enzyme-linked immunosorbent assay (ELISA), and comparative genomic hybridization (CGH). Japanese breast cancer patients displayed significantly longer alleles for the CA-SSR I repeat ($p < 0.001$), associated with significantly lower EGFR expression (mean 65 versus 36 fmol/mg membrane protein). Allelic imbalance (restricted to CA-SSR I) was observed in 55% of the informative Japanese breast cancers compared with only 34% of the German breast cancer reference group. Using a quantitative 5′ nuclease assay for *egfr*, a significantly higher percentage of Japanese breast cancer patients revealed amplifications of the CA-SSR I repeat ($p < 0.01$). Japanese patients with these amplifications were characterized by a significantly higher EGFR content compared with the German breast cancer patients ($p < 0.05$). These data show, on the one hand, that the correlation of EGFR overexpression and an inherited CA repeat polymorphism within intron 1 of *egfr* is a general finding in breast cancer, as has been shown previously. On the other hand, the data demonstrate clearly for the first time an interaction between the length of a polymorphism in intron 1 of *egfr* as an inherited genetic factor and the frequency of *egfr* amplification, as an acquired genetic factor, both factors contributing to EGFR overexpression in breast cancer. This new knowledge about mechanisms of regulation of EGFR expression might serve as an additional basis for evaluating anti-EGFR-based therapies.

Keywords: **epidermal growth factor receptor; polymorphism; epidemiology**

Introduction

The epidermal growth factor receptor (EGFR) belongs to the *erb*B gene family and is overexpressed in a great variety of malignant tumours. In a number of studies investigating invasive breast cancer patients, overexpression of EGFR was found to correlate inversely with oestrogen receptor expression levels and positively with clinical outcome. Nevertheless, it was regarded as a prognostic factor of undetermined significance [1–3]. The fact that its clinical value has not to date been estimated might be because the regulation of EGFR in cancer is not yet adequately understood [4].

In a previously published study, we investigated a polymorphic CA repeat in intron 1 of the epidermal growth factor receptor gene (*egfr* CA-SSR I), which determines the basal transcription activity of the *egfr* gene [1,5]. Moreover, we observed that allelic imbalances (AIs) restricted to the CA-SSR

I, as detectable by PCR-based microsatellite analysis, are a frequent event in breast cancer. They were even detectable in non-malignant breast tissue adjacent to invasive and *in situ* breast cancer [6]. Using gene dosage PCR, we determined that such AIs represented amplifications involving the regulatory sequence of intron 1. Additionally, inter-ethnic differences in the length of the CA dinucleotide repeat have been reported [7] and length differences between sporadic and familial breast cancer patients [8] have been described. Preliminary clinical observations also point towards a possible value of this polymorphism in the prediction of chemosensitivity [9,10].

Improved understanding of gene transcription and expression of *egfr* is required for clinical purposes, in view of the development, and clinical study, of new synthetic and antibody-based anti-EGFR therapies in recent years. These therapeutic approaches rely on optimal pre-therapeutic assessment of individual *egfr* status. A variety of technical approaches, including ELISA techniques and immunohistochemistry, as well as gene dosage measurements, have been published in the past. Only a limited number of somewhat contradictory studies have been published focusing on *egfr* amplifications; these have demonstrated that high-level amplifications of the whole gene are very rare events in breast cancer, in sharp contrast to its sister gene, c-*erb*B2.

In order to provide insights into the relationship between the length of the CA-SSR I in intron 1 of *egfr* and EGFR expression, we investigated breast cancer tissue from Japanese and German breast cancer patients. These features differ significantly between these ethnic subgroups, indicating that new diagnostic and predictive algorithms may need to be developed in the future.

Materials and methods

Tumour and lymphocyte DNAs were extracted and used for further analysis after informed consent. The study protocol for the storage, analysis, and statistical evaluation of the samples was approved by the local ethical committees in Muenster, Nagoya, and Saitama.

Determination of the number of CA repeats in intron I

The germ line allele status of the polymorphic CA repeat within intron 1 of *egfr*, being a fragment of 114–138 bp (by PCR) containing the CA dinucleotide repeat, was determined as described elsewhere in 180 German and 126 Japanese breast cancer patients [1]. Separation was achieved using a four-colour laser-induced fluorescence capillary electrophoresis system (Prism 3700; ABI, Weiterstadt, Germany) employing GeneScan Standard TAMRA-500 for fragment length evaluation. Evaluation was performed using GeneScan™ 2.1 evaluation software (ABI, Weiterstadt, Germany).

Determination of loss of heterozygosity (LOH) in intron I

Tumour DNA and DNA from blood lymphocytes were isolated with the QIAamp Tissue Kit (Quiagen, Hilden, Germany) from 43 Japanese breast cancer patients. A 114–132 bp PCR fragment containing the polymorphic region was amplified with 50 pmol of established primers [11] and evaluated for allele scoring and assessment of LOH as described elsewhere [1,12].

Determination of EGFR expression (ELISA)

EGFR content in breast cancer samples was analysed in 30 Japanese and 30 German tumour samples. The samples were pair-matched by menopausal status, histological grade, and oestrogen receptor status. All samples were analysed in the same run. Protein isolation and the ELISA for the determination of EGFR expression were carried out as described elsewhere [1].

Quantitative real-time PCR (5′ nuclease assay)

Primers specific for sequences flanking the first CA repeat in the first intron of the *egfr* gene (CAIfor: 5′-TGAAGAATTTGAGCCAACCAAA-3′ and CAIrev: 5′-CACTTGAACCAGGGACAGCA-3′) were designed using Primer Express software (Applera), and a universal probe consisting of 15 CA repeats as well as a 5′ fluorescent label (CA-Fam) was also established. Primers and probes were also designed for two different single-copy genes, manganese superoxide dismutase (SOD2, chromosome 6q25, GenBank accession No 65 965) and beta-globin (HBB, chromosome 11p, GenBank accession No V00 499). The 5′ nuclease assay including the appropriate controls was performed as previously described [6]. DNA concentrations were normalized with respect to both SOD2 and HBB, two different single-copy genes [13].

Statistical tests

Statistical analysis was carried out using the Wilcoxon rank test, Student's *t*-test and the chi-square-test. p values less than 0.05 were considered significant.

Results

Comparison of the frequency of allelic imbalance (AI) at CA-SSR I as determined by PCR-based microsatellite analysis

Sixty-two per cent of the Japanese breast tumours were heterozygous (informative) for the microsatellite analysed. As such, 55% of these tumours revealed AI at CA-SSR I. However, in the previously reported German comparator group, 88% of tumours were heterozygous but AI at CA-SSR I was present in only 34% ($p < 0.01$) [1,6].