1 Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet.* **365,** 1687 1717 (2005).
2 Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* **351,** 1451 1467 (1998).
3 Wakeling, A. E. & Slater, S. R. Estrogen-receptor binding and biologic activity of tamoxifen and its metabolites. *Cancer Treat. Rep.* **64,** 741 744 (1980).
4 Murdter, T. E., Schroth, W., Bacchus-Gerybadze, L., Winter, S., Heinkele, G., Simon, W. *et al.* Activity levels of tamoxifen metabolites at the estrogen receptor and the impact of genetic polymorphisms of phase I and II enzymes on their concentration levels in plasma. *Clin. Pharmacol. Ther.* **89,** 708 717 (2011).
5 Boocock, D. J., Brown, K., Gibbs, A. H., Sanchez, E., Turteltaub, K. W. & White, I. N. Identification of human CYP forms involved in the activation of tamoxifen and irreversible binding to DNA. *Carcinogenesis* **23,** 1897 1901 (2002).
6 Crewe, H. K., Notley, L. M., Wunsch, R. M., Lennard, M. S. & Gillam, E. M. Metabolism of tamoxifen by recombinant human cytochrome P450 enzymes: formation of the 4-hydroxy, 4'-hydroxy and *N*-desmethyl metabolites and isomerization of *trans*-4-hydroxytamoxifen. *Drug Metab. Dispos.* **30,** 869 874 (2002).
7 Coller, J. K., Krebsfaenger, N., Klein, K., Wolbold, R., Nussler, A., Neuhaus, P. *et al.* Large interindividual variability in the *in vitro* formation of tamoxifen metabolites related to the development of genotoxicity. *Br. J. Clin. Pharmacol.* **57,** 105 111 (2004).
8 Borges, S., Desta, Z., Li, L., Skaar, T. C., Ward, B. A., Nguyen, A. *et al.* Quantitative effect of *CYP2D6* genotype and inhibitors on tamoxifen metabolism: implication for optimization of breast cancer treatment. *Clin. Pharmacol. Ther.* **80,** 61 74 (2006).
9 Jin, Y., Desta, Z., Stearns, V., Ward, B., Ho, H., Lee, K. H. *et al.* CYP2D6 genotype, antidepressant use, and tamoxifen metabolism during adjuvant breast cancer treatment. *J. Natl. Cancer Inst.* **97,** 30 39 (2005).
10 Johnson, M. D., Zuo, H., Lee, K. H., Trebley, J. P., Rae, J. M., Weatherman, R. V. *et al.* Pharmacological characterization of 4-hydroxy-*N*-desmethyl tamoxifen, a novel active metabolite of tamoxifen. *Breast Cancer Res. Treat.* **85,** 151 159 (2004).
11 Stearns, V., Johnson, M. D., Rae, J. M., Morocho, A., Novielli, A., Bhargava, P. *et al.* Active tamoxifen metabolite plasma concentrations after coadministration of tamoxifen and the selective serotonin reuptake inhibitor paroxetine. *J. Natl. Cancer Inst.* **95,** 1758 1764 (2003).
12 Clarke, R., Liu, M. C., Bouker, K. B., Gu, Z., Lee, R. Y., Zhu, Y. *et al.* Antiestrogen resistance in breast cancer and the role of estrogen receptor signaling. *Oncogene* **22,** 7316 7339 (2003).
13 Crewe, H. K., Ellis, S. W., Lennard, M. S. & Tucker, G. T. Variable contribution of cytochromes P450 2D6, 2C9 and 3A4 to the 4-hydroxylation of tamoxifen by human liver microsomes. *Biochem. Pharmacol.* **53,** 171 178 (1997).
14 Dehal, S. S. & Kupfer, D. CYP2D6 catalyzes tamoxifen 4-hydroxylation in human liver. *Cancer Res.* **57,** 3402 3406 (1997).
15 Coller, J. K., Krebsfaenger, N., Klein, K., Endrizzi, K., Wolbold, R., Lang, T. *et al.* The influence of *CYP2B6*, *CYP2C9* and *CYP2D6* genotypes on the formation of the potent antioestrogen Z-4-hydroxy-tamoxifen in human liver. *Br. J. Clin. Pharmacol.* **54,** 157 167 (2002).
16 Borgna, J. L. & Rochefort, H. Hydroxylated metabolites of tamoxifen are formed *in vivo* and bound to estrogen receptor in target tissues. *J. Biol. Chem.* **256,** 859 868 (1981).
17 Lien, E. A., Solheim, E., Lea, O. A., Lundgren, S., Kvinnsland, S. & Ueland, P. M. Distribution of 4-hydroxy-*N*-desmethyltamoxifen and other tamoxifen metabolites in human biological fluids during tamoxifen treatment. *Cancer Res.* **49,** 2175 2183 (1989).
18 Desta, Z., Ward, B. A., Soukhova, N. V. & Flockhart, D. A. Comprehensive evaluation of tamoxifen sequential biotransformation by the human cytochrome P450 system *in vitro*: prominent roles for CYP3A and CYP2D6. *J. Pharmacol. Exp. Ther.* **310,** 1062 1075 (2004).
19 Falany, J. L., Pilloff, D. E., Leyh, T. S. & Falany, C. N. Sulfation of raloxifene and 4-hydroxytamoxifen by human cytosolic sulfotransferases. *Drug Metab. Dispos.* **34,** 361 368 (2006).
20 Gjerde, J., Hauglid, M., Breilid, H., Lundgren, S., Varhaug, J. E., Kisanga, E. R. *et al.* Effects of CYP2D6 and SULT1A1 genotypes including SULT1A1 gene copy number on tamoxifen metabolism. *Ann. Oncol.* **19,** 56 61 (2008).
21 Nishiyama, T., Ogura, K., Nakano, H., Ohnuma, T., Kaku, T., Hiratsuka, A. *et al.* Reverse geometrical selectivity in glucuronidation and sulfation of *cis*- and *trans*-4-hydroxytamoxifens by human liver UDP-glucuronosyltransferases and sulfotransferases. *Biochem. Pharmacol.* **63,** 1817 1830 (2002).
22 Ogura, K., Ishikawa, Y., Kaku, T., Nishiyama, T., Ohnuma, T., Muro, K. *et al.* Quaternary ammonium-linked glucuronidation of *trans*-4-hydroxytamoxifen, an active metabolite of tamoxifen, by human liver microsomes and UDP-glucuronosyltransferase 1A4. *Biochem. Pharmacol.* **71,** 1358 1369 (2006).
23 Sun, D., Sharma, A. K., Dellinger, R. W., Blevins-Primeau, A. S., Balliet, R. M., Chen, G. *et al.* Glucuronidation of active tamoxifen metabolites by the human UDP glucuronosyltransferases. *Drug Metab. Dispos.* **35,** 2006 2014 (2007).
24 Kaku, T., Ogura, K., Nishiyama, T., Ohnuma, T., Muro, K. & Hiratsuka, A. Quaternary ammonium-linked glucuronidation of tamoxifen by human liver microsomes and UDP-glucuronosyltransferase 1A4. *Biochem. Pharmacol.* **67,** 2093 2102 (2004).
25 Sun, D., Chen, G., Dellinger, R. W., Duncan, K., Fang, J. L. & Lazarus, P. Characterization of tamoxifen and 4-hydroxytamoxifen glucuronidation by human UGT1A4 variants. *Breast Cancer Res.* **8,** R50 (2006).
26 Ingelman-Sundberg, M. Genetic polymorphisms of cytochrome *P*450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity. *Pharmacogenomics J.* **5,** 6 13 (2005).
27 Bradford, L. D. *CYP2D6* allele frequency in European Caucasians, Asians, Africans and their descendants. *Pharmacogenomics* **3,** 229 243 (2002).
28 Sachse, C., Brockmoller, J., Bauer, S. & Roots, I. Cytochrome P450 2D6 variants in a Caucasian population: allele frequencies and phenotypic consequences. *Am. J. Hum. Genet.* **60,** 284 295 (1997).
29 Griese, E. U., Zanger, U. M., Brudermanns, U., Gaedigk, A., Mikus, G., Morike, K. *et al.* Assessment of the predictive power of genotypes for the *in-vivo* catalytic function of CYP2D6 in a German population. *Pharmacogenetics* **8,** 15 26 (1998).
30 Broly, F., Gaedigk, A., Heim, M., Eichelbaum, M., Morike, K. & Meyer, U. A. Debrisoquine/sparteine hydroxylation genotype and phenotype: analysis of common mutations and alleles of CYP2D6 in a European population. *DNA Cell Biol.* **10,** 545 558 (1991).
31 Sachse, C., Brockmoller, J., Hildebrand, M., Muller, K. & Roots, I. Correctness of prediction of the CYP2D6 phenotype confirmed by genotyping 47 intermediate and poor metabolizers of debrisoquine. *Pharmacogenetics* **8,** 181 185 (1998).
32 Nakamura, K., Goto, F., Ray, W. A., McAllister, C. B., Jacqz, E., Wilkinson, G. R. *et al.* Interethnic differences in genetic polymorphism of debrisoquin and mephenytoin hydroxylation between Japanese and Caucasian populations. *Clin. Pharmacol. Ther.* **38,** 402 408 (1985).
33 Yokota, H., Tamura, S., Furuya, H., Kimura, S., Watanabe, M., Kanazawa, I. *et al.* Evidence for a new variant *CYP2D6* allele *CYP2D6J* in a Japanese population associated with lower *in vivo* rates of sparteine metabolism. *Pharmacogenetics* **3,** 256 263 (1993).
34 Tateishi, T., Chida, M., Ariyoshi, N., Mizorogi, Y., Kamataki, T. & Kobayashi, S. Analysis of the *CYP2D6* gene in relation to dextromethorphan *O*-demethylation capacity in a Japanese population. *Clin. Pharmacol. Ther.* **65,** 570 575 (1999).
35 Wang, S. L., Huang, J. D., Lai, M. D., Liu, B. H. & Lai, M. L. Molecular basis of genetic variation in debrisoquin hydroxylation in Chinese subjects: polymorphism in RFLP and DNA sequence of *CYP2D6*. *Clin. Pharmacol. Ther.* **53,** 410 418 (1993).
36 Yokoi, T., Kosaka, Y., Chida, M., Chiba, K., Nakamura, H., Ishizaki, T. *et al.* A new *CYP2D6* allele with a nine base insertion in exon 9 in a Japanese population associated with poor metabolizer phenotype. *Pharmacogenetics* **6,** 395 401 (1996).
37 Chida, M., Yokoi, T., Nemoto, N., Inaba, M., Kinoshita, M. & Kamataki, T. A new variant *CYP2D6* allele (*CYP2D6*21*) with a single base insertion in exon 5 in a Japanese population associated with a poor metabolizer phenotype. *Pharmacogenetics* **9,** 287 293 (1999).
38 Yamazaki, H., Kiyotani, K., Tsubuko, S., Matsunaga, M., Fujieda, M., Saito, T. *et al.* Two novel haplotype of *CYP2D6* gene in a Japanese population. *Drug Metab. Pharmacokinet.* **18,** 269 271 (2003).
39 Lim, H. S., Lee, J. H., Lee, S. K., Lee, S. E., Jang, I. J. & Ro, J. Clinical implications of CYP2D6 genotypes predictive of tamoxifen pharmacokinetics in metastatic breast cancer. *J. Clin. Oncol.* **25,** 3837 3845 (2007).
40 Kiyotani, K., Mushiroda, T., Imamura, C. K., Hosono, N., Tsunoda, T., Kubo, M. *et al.* Significant effect of polymorphisms in *CYP2D6* and *ABCC2* on clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients. *J. Clin. Oncol.* **28,** 1287 1293 (2010).
41 Lim, J. S., Chen, X. A., Singh, O., Yap, Y. S., Ng, R. C., Wong, N. S. *et al.* Impact of *CYP2D6*, *CYP3A5*, *CYP2C9* and *CYP2C19* polymorphisms on tamoxifen pharmacokinetics in Asian breast cancer patients. *Br. J. Clin. Pharmacol.* **71,** 737 750 (2011).
42 Goetz, M. P., Rae, J. M., Suman, V. J., Safgren, S. L., Ames, M. M., Visscher, D. W. *et al.* Pharmacogenetics of tamoxifen biotransformation is associated with clinical outcomes of efficacy and hot flashes. *J. Clin. Oncol.* **23,** 9312 9318 (2005).
43 Goetz, M. P., Knox, S. K., Suman, V. J., Rae, J. M., Safgren, S. L., Ames, M. M. *et al.* The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. *Breast Cancer Res. Treat.* **101,** 113 121 (2007).
44 Schroth, W., Goetz, M. P., Hamann, U., Fasching, P. A., Schmidt, M., Winter, S. *et al.* Association between CYP2D6 polymorphisms and outcomes among women with early stage breast cancer treated with tamoxifen. *JAMA* **302,** 1429 1436 (2009).
45 Schroth, W., Antoniadou, L., Fritz, P., Schwab, M., Muerdter, T., Zanger, U. M. *et al.* Breast cancer treatment outcome with adjuvant tamoxifen relative to patient CYP2D6 and CYP2C19 genotypes. *J. Clin. Oncol.* **25,** 5187 5193 (2007).
46 Newman, W. G., Hadfield, K. D., Latif, A., Roberts, S. A., Shenton, A., McHague, C. *et al.* Impaired tamoxifen metabolism reduces survival in familial breast cancer patients. *Clin. Cancer Res.* **14,** 5913 5918 (2008).
47 Bijl, M. J., van Schaik, R. H., Lammers, L. A., Hofman, A., Vulto, A. G., van Gelder, T. *et al.* The CYP2D6*4 polymorphism affects breast cancer survival in tamoxifen users. *Breast Cancer Res. Treat.* **118,** 125 130 (2009).
48 Ramón, Y., Altés, A., Paré, L., del Rio, E., Alonso, C., Barnadas, A. *et al.* Impact of CYP2D6 polymorphisms in tamoxifen adjuvant breast cancer treatment. *Breast Cancer Res. Treat.* **119,** 33 38 (2010).
49 Damodaran, S. E., Pradhan, S. C., Umamaheswaran, G., Kadambari, D., Reddy, K. S. & Adithan, C. Genetic polymorphisms of *CYP2D6* increase the risk for recurrence of breast cancer in patients receiving tamoxifen as an adjuvant therapy. *Cancer Chemother. Pharmacol.* **70,** 75 81 (2012).
50 Goetz, M. P., Suman, V. J., Hoskin, T. L., Gnant, M., Filipits, M., Safgren, S. L. *et al.* CYP2D6 metabolism and patient outcome in the Austrian Breast and Colorectal Cancer Study Group trial (ABCSG) 8. *Clin. Cancer Res.* **19,** 500 507 (2013).
51 Kiyotani, K., Mushiroda, T., Sasa, M., Bando, Y., Sumitomo, I., Hosono, N. *et al.* Impact of *CYP2D6*10* on recurrence-free survival in breast cancer patients receiving adjuvant tamoxifen therapy. *Cancer Sci.* **99,** 995 999 (2008).

52 Xu, Y., Sun, Y., Yao, L., Shi, L., Wu, Y., Ouyang, T. *et al.* Association between CYP2D6*10 genotype and survival of breast cancer patients receiving tamoxifen treatment. *Ann. Oncol.* **19,** 1423 1429 (2008).
53 Park, H. S., Choi, J. Y., Lee, M. J., Park, S., Yeo, C. W., Lee, S. S. *et al.* Association between genetic polymorphisms of *CYP2D6* and outcomes in breast cancer patients with tamoxifen treatment. *J. Korean Med. Sci.* **26,** 1007 1013 (2011).
54 Teh, L. K., Mohamed, N. I., Salleh, M. Z., Rohaizak, M., Shahrun, N. S., Saladina, J. J. *et al.* The risk of recurrence in breast cancer patients treated with tamoxifen: polymorphisms of *CYP2D6* and *ABCB1*. *AAPS J.* **14,** 52 59 (2012).
55 Sukasem, C., Sirachainan, E., Chamnanphon, M., Pechatanan, K., Sirisinha, T., Ativitavas, T. *et al.* Impact of *CYP2D6* polymorphisms on tamoxifen responses of women with breast cancer: a microarray-based study in Thailand. *Asian Pac. J. Cancer Prev.* **13,** 4549 4553 (2012).
56 Nowell, S. A., Ahn, J., Rae, J. M., Scheys, J. O., Trovato, A., Sweeney, C. *et al.* Association of genetic variation in tamoxifen-metabolizing enzymes with overall survival and recurrence of disease in breast cancer patients. *Breast Cancer Res. Treat.* **91,** 249 258 (2005).
57 Wegman, P., Vainikka, L., Stal, O., Nordenskjold, B., Skoog, L., Rutqvist, L. E. *et al.* Genotype of metabolic enzymes and the benefit of tamoxifen in postmenopausal breast cancer patients. *Breast Cancer Res.* **7,** R284 R290 (2005).
58 Wegman, P., Elingarami, S., Carstensen, J., Stal, O., Nordenskjold, B. & Wingren, S. Genetic variants of *CYP3A5*, *CYP2D6*, *SULT1A1*, *UGT2B15* and tamoxifen response in postmenopausal patients with breast cancer. *Breast Cancer Res.* **9,** R7 (2007).
59 Okishiro, M., Taguchi, T., Jin Kim, S., Shimazu, K., Tamaki, Y. & Noguchi, S. Genetic polymorphisms of CYP2D6*10 and CYP2C19*2, *3 are not associated with prognosis, endometrial thickness, or bone mineral density in Japanese breast cancer patients treated with adjuvant tamoxifen. *Cancer* **115,** 952 961 (2009).
60 Abraham, J. E., Maranian, M. J., Driver, K. E., Platte, R., Kalmyrzaev, B., Baynes, C. *et al.* CYP2D6 gene variants: association with breast cancer specific survival in a cohort of breast cancer patients from the United Kingdom treated with adjuvant tamoxifen. *Breast Cancer Res.* **12,** R64 (2010).
61 Park, I. H., Ro, J., Park, S., Lim, H. S., Lee, K. S., Kang, H. S. *et al.* Lack of any association between functionally significant *CYP2D6* polymorphisms and clinical outcomes in early breast cancer patients receiving adjuvant tamoxifen treatment. *Breast Cancer Res. Treat.* **131,** 455 461 (2012).
62 Rae, J. M., Drury, S., Hayes, D. F., Stearns, V., Thibert, J. N., Haynes, B. P. *et al.* *CYP2D6* and *UGT2B7* genotype and risk of recurrence in tamoxifen-treated breast cancer patients. *J. Natl Cancer Inst.* **104,** 452 460 (2012).
63 Regan, M. M., Leyland-Jones, B., Bouzyk, M., Pagani, O., Tang, W., Kammler, R. *et al.* *CYP2D6* genotype and tamoxifen response in postmenopausal women with endocrine-responsive breast cancer: the breast international group 1-98 trial. *J. Natl Cancer Inst.* **104,** 441 451 (2012).
64 Pharoah, P. D., Abraham, J. & Caldas, C. Re: *CYP2D6* genotype and tamoxifen response in postmenopausal women with endocrine-responsive breast cancer: the Breast International Group 1-98 trial and Re: *CYP2D6* and *UGT2B7* genotype and risk of recurrence in tamoxifen-treated breast cancer patients. *J. Natl Cancer Inst.* **104,** 1263 1264 (2012).
65 Nakamura, Y., Ratain, M. J., Cox, N. J., McLeod, H. L., Kroetz, D. L. & Flockhart, D. A. Re: CYP2D6 genotype and tamoxifen response in postmenopausal women with endocrine-responsive breast cancer: the Breast International Group 1-98 trial. *J. Natl Cancer Inst.* **104,** 1264 (2012).
66 Stanton Jr V. Re: CYP2D6 genotype and tamoxifen response in postmenopausal women with endocrine-responsive breast cancer: the Breast International Group 1-98 trial. *J. Natl Cancer Inst* **104,** 1265 1266 (2012).
67 Brauch, H., Schroth, W., Goetz, M. P., Murdter, T. E., Winter, S., Ingle, J. N. *et al.* Tamoxifen use in postmenopausal breast cancer: CYP2D6 matters. *J. Clin. Oncol.* **31,** 176 180 (2013).
68 Schroth, W., Hamann, U., Fasching, P. A., Dauser, S., Winter, S., Eichelbaum, M. *et al.* *CYP2D6* polymorphisms as predictors of outcome in breast cancer patients treated with tamoxifen: expanded polymorphism coverage improves risk stratification. *Clin. Cancer Res.* **16,** 4468 4477 (2010).
69 Hirano, A., Emi, M., Tsuneizumi, M., Utada, Y., Yoshimoto, M., Kasumi, F. *et al.* Allelic losses of loci at 3p25.1, 8p22, 13q12, 17p13.3, and 22q13 correlate with postoperative recurrence in breast cancer. *Clin. Cancer Res.* **7,** 876 882 (2001).
70 Kiyotani, K., Mushiroda, T., Hosono, N., Tsunoda, T., Kubo, M., Aki, F. *et al.* Lessons for pharmacogenomics studies: association study between *CYP2D6* genotype and tamoxifen response. *Pharmacogenet. Genomics* **20,** 565 568 (2010).
71 Kiyotani, K., Mushiroda, T., Nakamura, Y. & Zembutsu, H. Pharmacogenomics of tamoxifen: roles of drug metabolizing enzymes and transporters. *Drug Metab. Pharmacokinet.* **27,** 122 131 (2012).
72 Madlensky, L., Natarajan, L., Tchu, S., Pu, M., Mortimer, J., Flatt, S. W. *et al.* Tamoxifen metabolite concentrations, *CYP2D6* genotype, and breast cancer outcomes. *Clin. Pharmacol. Ther.* **89,** 718 725 (2011).
73 Barginear, M. F., Jaremko, M., Peter, I., Yu, C., Kasai, Y., Kemeny, M. *et al.* Increasing tamoxifen dose in breast cancer patients based on CYP2D6 genotypes and endoxifen levels: effect on active metabolite isomers and the antiestrogenic activity score. *Clin. Pharmacol. Ther.* **90,** 605 611 (2011).
74 Irvin, W. J. Jr., Walko, C. M., Weck, K. E., Ibrahim, J. G., Chiu, W. K., Dees, E. C. *et al.* Genotype-guided tamoxifen dosing increases active metabolite exposure in women with reduced CYP2D6 metabolism: a multicenter study. *J. Clin. Oncol.* **29,** 3232 3239 (2011).
75 Kiyotani, K., Mushiroda, T., Imamura, C. K., Tanigawara, Y., Hosono, N., Kubo, M. *et al.* Dose-adjustment study of tamoxifen based on *CYP2D6* genotypes in Japanese breast cancer patients. *Breast Cancer Res. Treat.* **131,** 137 145 (2012).
76 Tucker, A. N., Tkaczuk, K. A., Lewis, L. M., Tomic, D., Lim, C. K. & Flaws, J. A. Polymorphisms in cytochrome P4503A5 (CYP3A5) may be associated with race and tumor characteristics, but not metabolism and side effects of tamoxifen in breast cancer patients. *Cancer Lett.* **217,** 61 72 (2005).
77 Gjerde, J., Geisler, J., Lundgren, S., Ekse, D., Varhaug, J. E., Mellgren, G. *et al.* Associations between tamoxifen, estrogens, and FSH serum levels during steady state tamoxifen treatment of postmenopausal women with breast cancer. *BMC Cancer* **10,** 313 (2010).
78 Ruiter, R., Bijl, M. J., van Schaik, R. H., Berns, E. M., Hofman, A., Coebergh, J. W. *et al.* *CYP2C19*2* polymorphism is associated with increased survival in breast cancer patients using tamoxifen. *Pharmacogenomics* **11,** 1367 1375 (2010).
79 Moyer, A. M., Suman, V. J., Weinshilboum, R. M., Avula, R., Black, J. L., Safgren, S. L. *et al.* *SULT1A1*, *CYP2C19* and disease-free survival in early breast cancer patients receiving tamoxifen. *Pharmacogenomics* **12,** 1535 1543 (2011).
80 Hosono, N., Kato, M., Kiyotani, K., Mushiroda, T., Takata, S., Sato, H. *et al.* *CYP2D6* genotyping for functional-gene dosage analysis by allele copy number detection. *Clin. Chem.* **55,** 1546 1554 (2009).
81 Hosono, N., Kubo, M., Tsuchiya, Y., Sato, H., Kitamoto, T., Saito, S. *et al.* Multiplex PCR-based real-time invader assay (mPCR-RETINA): a novel SNP-based method for detecting allelic asymmetries within copy number variation regions. *Hum. Mutat.* **29,** 182 189 (2008).
82 Kiyotani, K., Shimizu, M., Kumai, T., Kamataki, T., Kobayashi, S. & Yamazaki, H. Limited effects of frequent *CYP2D6*36-*10* tandem duplication allele on *in vivo* dextromethorphan metabolism in a Japanese population. *Eur. J. Clin. Pharmacol.* **66,** 1065 1068 (2010).
83 Goda, K., Bacso, Z. & Szabo, G. Multidrug resistance through the spectacle of P-glycoprotein. *Curr. Cancer Drug Targets* **9,** 281 297 (2009).
84 Teft, W. A., Mansell, S. E. & Kim, R. B. Endoxifen, the active metabolite of tamoxifen, is a substrate of the efflux transporter P-glycoprotein (multidrug resistance 1). *Drug Metab. Dispos.* **39,** 558 562 (2011).
85 Iusuf, D., Teunissen, S. F., Wagenaar, E., Rosing, H., Beijnen, J. H. & Schinkel, A. H. P-glycoprotein (ABCB1) transports the primary active tamoxifen metabolites endoxifen and 4-hydroxytamoxifen and restricts their brain penetration. *J. Pharmacol. Exp. Ther.* **337,** 710 717 (2011).
86 Kiyotani, K., Mushiroda, T., Nakamura, Y. & Zembutsu, H. *ABCC2* and clinical outcome of tamoxifen therapy: reply to T. Lang et al. *J. Clin. Oncol.* **28,** e449 (2010).
87 Choi, H. K., Yang, J. W., Roh, S. H., Han, C. Y. & Kang, K. W. Induction of multidrug resistance associated protein 2 in tamoxifen-resistant breast cancer cells. *Endocr. Relat. Cancer* **14,** 293 303 (2007).
88 Kiyotani, K., Mushiroda, T., Tsunoda, T., Morizono, T., Hosono, N., Kubo, M. *et al.* A genome-wide association study identifies locus at 10q22 associated with clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients in Japanese. *Hum. Mol. Genet.* **21,** 1665 1672 (2012).
89 Musgrove, E. A. & Sutherland, R. L. Biological determinants of endocrine resistance in breast cancer. *Nat. Rev. Cancer* **9,** 631 643 (2009).
90 Giuliano, M., Schifp, R., Osborne, C. K. & Trivedi, M. V. Biological mechanisms and clinical implications of endocrine resistance in breast cancer. *Breast* **20** (Suppl 3), S42 S49 (2011).
91 Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA* **98,** 10869 10874 (2001).
92 Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl Acad. Sci. USA* **100,** 8418 8423 (2003).
93 Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M. *et al.* A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351,** 2817 2826 (2004).
94 Ma, X. J., Wang, Z., Ryan, P. D., Isakoff, S. J., Barmettler, A., Fuller, A. *et al.* A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* **5,** 607 616 (2004).
95 Jansen, M. P., Sieuwerts, A. M., Look, M. P., Ritstier, K., Meijer-van Gelder, M. E., van Staveren, I. L. *et al.* *HOXB13*-to-*IL17BR* expression ratio is related with tumor aggressiveness and response to tamoxifen of recurrent breast cancer: a retrospective study. *J. Clin. Oncol.* **25,** 662 668 (2007).
96 Goetz, M. P., Suman, V. J., Couch, F. J., Ames, M. M., Rae, J. M., Erlander, M. G. *et al.* Cytochrome P450 2D6 and homeobox 13/interleukin-17B receptor: combining inherited and tumor gene markers for prediction of tamoxifen resistance. *Clin. Cancer Res.* **14,** 5864 5868 (2008).
97 Ellis, M. J., Ding, L., Shen, D., Luo, J., Suman, V. J., Wallis, J. W. *et al.* Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* **486,** 353 360 (2012).

Cancer Science
The official journal of the Japanese Cancer Association

Genome-wide association study of chemotherapeutic agent-induced severe neutropenia/leucopenia for patients in Biobank Japan

Siew-Kee Low,[1,4] Suyoun Chung,[4,5] Atsushi Takahashi,[1] Hitoshi Zembutsu,[4] Taisei Mushiroda,[2] Michiaki Kubo[3] and Yusuke Nakamura[4,5,6]

Laboratories for [1]Statistical Analysis; [2]Pharmacogenetics; [3]Genotyping Development, Center for Genomic Medicine, RIKEN, Yokohama; [4]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; [5]Department of Medicine, The University of Chicago, Chicago, Illinois, USA

(Received March 21, 2013/Revised April 22, 2013/Accepted April 22, 2013/Accepted manuscript online May 4, 2013/Article first published online June 10, 2013)

Chemotherapeutic agents are notoriously known to have a narrow therapeutic range that often results in life-threatening toxicity. Hence, it is clinically important to identify the patients who are at high risk for severe toxicity to certain chemotherapy through a pharmacogenomics approach. In this study, we carried out multiple genome-wide association studies (GWAS) of 13 122 cancer patients who received different chemotherapy regimens, including cyclophosphamide- and platinum-based (cisplatin and carboplatin), anthracycline-based (doxorubicin and epirubicin), and antimetabolite-based (5-fluorouracil and gemcitabine) treatment, antimicrotubule agents (paclitaxel and docetaxel), and topoisomerase inhibitors (camptothecin and etoposide), as well as combination therapy with paclitaxel and carboplatin, to identify genetic variants that are associated with the risk of severe neutropenia/leucopenia in the Japanese population. In addition, we used a weighted genetic risk scoring system to evaluate the cumulative effects of the suggestive genetic variants identified from GWAS in order to predict the risk levels of individuals who carry multiple risk alleles. Although we failed to identify genetic variants that surpassed the genome-wide significance level ($P < 5.0 \times 10^{-8}$) through GWAS, probably due to insufficient statistical power and complex clinical features, we were able to shortlist some of the suggestive associated loci. The current study is at the relatively preliminary stage, but does highlight the complexity and problematic issues associated with retrospective pharmacogenomics studies. However, we hope that verification of these genetic variants through local and international collaborations could improve the clinical outcome for cancer patients. (*Cancer Sci* 2013; 104: 1074–1082)

It is now widely and well recognized that medication can cause distinct heterogeneity in terms of its efficacy and toxicity among individuals. These inter individual differences could be explained in part by the common and/or rare genetic variants in the human genome. Pharmacogenomics aims to discover how genetic variations in the human genome can affect a drug's efficacy or toxicity, and thus brings great promise for personalized medicine in which genetic information can be used to predict the safety, toxicity, and/or efficacy of drugs.[1] Pharmacogenomics study for chemotherapeutic therapies is particularly important because these drugs are known to have a narrow therapeutic window; in general, a higher concentration causes toxicity and a lower concentration reduces the efficacy of the drug. Two of the well described examples are the association of genetic variants in *TPMT* with 6 mercaptopurine induced myelosuppression in treatment of pediatric acute lymphoblastic leukemia and that of *UGT1A1* variants with camptothecin related neutropenia and diarrhea in treatment of colorectal and lung cancers. The US Food and Drug Administration have recommended that variants on these two genes should be helpful for the prediction of severe adverse reactions prior to use of the drugs.[2–7]

With advances in various technologies in the life sciences, it is now possible to accurately genotype more than a million common genetic variations by genome wide high density SNP array or to characterize all genetic variants in our genome by the next generation DNA sequencing methods. Although one of the greatest drawbacks of GWAS is the requirement of the large number of samples to achieve high statistical power,[8] this issue could be overcome by the establishment of Biobank Japan in 2003 (http://biobankjp.org/).[9] Biobank Japan collected approximately 330 000 disease cases (200 000 individuals) that had either one or multiples of 47 different diseases including cancers from a collaborative network of 66 hospitals throughout Japan, with the major aim to identify genetic variants associated with susceptibility to complex diseases or those related to drug toxicity. By using the samples from Biobank Japan, a significant number of insightful findings have been published in recent years for identification of common genetic variants associated with complex diseases including cancer.[10–19] With a reasonable number of samples, it is also feasible to carry out pharmacogenomics studies on chemotherapy induced toxicity.

Neutropenia and/or leucopenia are two of the most common drug adverse events after treatment with chemotherapeutic agents, which often cause life threatening infections and the delay of treatment schedule that subsequently affect the treatment outcome. Although prophylactic granulocyte colony stimulating factor has been given to the patients as a preventive measure,[20] the underlying mechanism and susceptible risk factors that cause neutropenia have not been fully elucidated. In this study, we carried out a total of 17 sets of GWAS using 13 122 cancer patients, who received various drug regimens, to identify genetic variants associated with the risk of chemotherapeutic agent induced severe neutropenia/leucopenia in the Japanese population.

Subjects and Method

Study subjects. A total of 13 122 DNA samples from cancer patients, who received various chemotherapeutic agents, stored in Biobank Japan (University of Tokyo, Tokyo, Japan), were used in this study. Among them, 805 patients developed severe neutropenia and/or leucopenia (≥grade 3), and 4804 patients

[6]To whom correspondence should be addressed.
E-mails: yusuke@ims.u-tokyo.ac.jp; ynakamura@bsd.uchicago.edu

doi: 10.1111/cas.12186
© 2013 Japanese Cancer Association

were not reported to develop any adverse reactions after being given chemotherapeutic agents. The samples could be classi fied into subgroups according to the drugs used: an alkylating agent (cyclophosphamide); platinum based (cisplatin and carboplatin), anthracycline based (doxorubicin and epirubicin); antimetabolite based (5 fluorouracil and gemcitabine), antimi crotubule based (paclitaxel and docetaxel); and topoisomerase inhibitor based (camptothecin and etoposide). The grade of toxicity was classified in accordance with the US National Cancer Institute's Common Toxicity Criteria version 2.0. The adverse event description is based on the medical records collected by the medical coordinator. The patients' demo graphic details are summarized in Table 1. Participants of this study provided written inform consent and this project was approved by the ethical committee from the Institute of Medical Sciences, University of Tokyo and the RIKEN Center for Genomic Medicine (Yokohama, Japan).

Genotyping and quality controls. DNAs obtained from the patients' blood were genotyped using Illumina OmniExpress BeadChip (San Diego, CA, USA) that contained 733 202 SNPs. Sample quality control was carried out by methods including identity by state to evaluate cryptic relatedness for each sample and population stratification by the use of princi pal component analysis to exclude genetically heterogeneous samples from further analysis.[21,22] Then, our standard SNP quality control was carried out by excluding SNPs deviating from the Hardy Weinberg equilibrium ($P \leq 1.0 \times 10^{-6}$), non polymorphic SNPs, SNPs with a call rate of <0.99, and those on the X chromosome.[21,22] Q Q plot and lambda values, which were calculated between observed P values from Fish er's test allelic model against expected P values, were used to further evaluate population substructure.

Statistical analysis. Genome wide case control association analyses were evaluated using Fisher's exact method consider ing allelic, dominant, and recessive genetic models. Manhattan plots of the study were generated using the minimum P value among the three genetic models for each SNP.

Scoring system using wGRS. The scoring analysis was carried out using SNPs with *Pmin* of $<1.0 \times 10^{-5}$ after exclusion of SNPs that are in strong linkage disequilibrium ($r^2 > 0.8$) in each GWAS. The wGRS were calculated according to De Jager *et al.*[23] Briefly, we first calculated the weight of each SNP that is the natural log of the odds ratio for each allele/genotype, con sidering the associated genetic model. For an additive model, we assigned a score of 2 to an individual with two risk alleles, 1 to that with one risk allele, and 0 to that with no risk allele. For a dominant model, we assigned a score of 1 to an individual with one or two risk alleles, and 0 to that with no risk allele. For a recessive model, we assigned a score of 1 to an individual with two risk alleles, and 0 to that with no or one risk allele. Then the cumulative genetic risk scores were determined by multiplying the number of risk alleles/genotype of each SNP by its corre sponding weight, and subsequently took the sum across the total number of SNPs that were taken into consideration of each GWAS set. We classified the genetics risk score into four differ ent groups created from the mean and SD: group 1, <mean − 1SD; group 2, mean − 1SD to mean; group 3, mean to mean + 1SD; and group 4, >mean + 1SD. Odds ratio, 95% confidence interval, P value, sensitivity, and specificity were calculated using group 1 as a reference. To calculate the OR in which one of the cells in the contingency table is zero, we applied the Haldane correction, used to avoid error in the calcu lation by adding 0.5 to all of the cells of a contingency table.

Results

After subdividing the patients by administered drugs/major drug subgroups, as previously mentioned, a total of 17 GWAS analyses were carried out by comparing the allele/genotype frequency between the patients who had developed severe neu tropenia/leucopenia (grade 3/4) to those who had not devel oped any adverse drug reactions. The Q Q plots of each GWAS and the calculated lambda value of below 1.00 indi cated no significant population stratification in each of these GWAS analyses (Fig. S1). From this study, although we could not identify any SNPs that surpassed the genome wide signifi cant threshold (P value $< 5 \times 10^{-8}$) for showing association with the risk of neutropenia/leucopenia induced by the certain type of drug or regimen, several possible candidate loci were identified. The results of the GWAS are summarized in Table 2, Table S1, and Figure S2; the results of wGRS are summarized in Table S2.

Table 1. Demographic details of cancer patients treated with chemotherapeutic agents, whose DNA samples are stored in Biobank Japan (The University of Tokyo, Tokyo, Japan)

Category	Controls†	Grade 1/2	Grade 3/4	Category	Controls†	Grade 1/2	Grade 3/4
All	4804	1253	805	Drug subtype			
Age, years (mean)	62.9	58.7	59.6	Alkylating agent	346	266	176
Gender				Cyclophosphamide	335	255	168
Male	2604	424	318	Platinum based	743	429	428
Female	2200	829	487	Cisplatin	471	191	176
Cancer subtype				Carboplatin	262	207	261
Lung cancer	587	259	266	Anthracycline	459	240	184
Breast cancer	876	388	204	Doxorubicin	66	85	83
Ovarian cancer	140	124	74	Epirubicin	370	132	83
Gastric cancer	827	100	56	Antimetabolite	2249	512	294
Esophageal cancer	208	65	53	5 Fluorouracil	952	331	177
Colorectal cancer	1573	161	50	Gemcitabine	226	111	80
Endometrial cancer	78	72	45	Antimicrotubule agent	825	468	371
Cervical cancer	129	57	35	Paclitaxel	364	321	218
Prostate cancer	91	13	21	Docetaxel	233	143	147
Pancreatic cancer	83	36	20	Topoisomerase inhibitor	187	123	106
Liver cancer	366	16	9	Camptothecin	155	106	59
Gallbladder cancer	56	9	1	Etoposide	39	19	54
				Paclitaxel + carboplatin	166	161	150

†Individuals who did not develop any adverse drug reactions after chemotherapy.

© 2013 Japanese Cancer Association

Table 2. Association analysis of single nucleotide polymorphisms (SNPs) with different chemotherapeutic drugs/drug subgroups known to induce severe neutropenia/leucopenia

CHR	SNP	BP	RA	NRA	RAF_Case	RAF_Ctr	P_a e c	P_dom	P_rec	Pm n	OR	L95	U95	*Gene*	re oc
Cyc ophospham de															
16	rs2519974*	22889186	T	C	0 503	0 381	2 52E-04	4 35E-06	2 77E-01	4 35E-06	1 647	1 264	2 146	*HS3ST2*	0
1	rs10922438*	198469162	T	C	0 214	0 106	6 01E-06	1 71E-05	7 10E-02	6 01E-06	2 301	1 608	3 293	*ATP6V1G3*	23190
19	rs3745571*	6475613	T	C	0 778	0 670	4 05E-04	7 72E-06	1 00E+00	7 72E-06	1 730	1 276	2 345	*DENND1C*	0
A p at num-based drugs															
15	rs4886670*	75449674	A	C	0 320	0 227	9 86E-07	1 43E-05	8 14E-04	9 86E-07	1 605	1 330	1 937	*RPL36AP45*	29318
19	rs33428*	30937843	G	A	0 481	0 403	2 71E-04	2 78E-06	3 11E-01	2 78E-06	1 375	1 160	1 629	*ZNF536*	0
14	rs12589282*	22937656	G	T	0 535	0 437	6 30E-06	5 42E-03	4 11E-06	4 11E-06	1 480	1 250	1 752	*TRA@*	0
3	rs3845905*	66525963	G	A	0 915	0 850	4 12E-06	7 45E-05	3 65E-04	4 12E-06	1 894	1 433	2 503	*LRIG1*	0
5	rs1895302*	169542600	C	T	0 551	0 478	6 95E-04	7 41E-06	4 11E-01	7 41E-06	1 340	1 132	1 587	*FOXI1*	5871
1	rs16825455*	21837755	T	C	0 686	0 605	8 85E-05	8 62E-06	1 90E-01	8 62E-06	1 425	1 193	1 702	*ALPL*	0
C sp at n															
7	rs10253216*	16861849	T	C	0 565	0 468	2 18E-03	1 68E-07	1 00E+00	1 68E-07	1 478	1 155	1 891	*AGR2*	−17111
4	rs11944965*	63424089	T	C	0 807	0 678	3 45E-06	1 68E-06	6 65E-02	1 68E-06	1 986	1 475	2 676	*LOC644534*	47600
7	rs7797977*	16862235	C	A	0 668	0 580	4 06E-03	5 23E-01	2 17E-06	2 17E-06	1 457	1 127	1 883	*AGR2*	−17497
18	rs2406342*	74488280	T	G	0 605	0 475	3 59E-05	2 48E-06	6 71E-02	2 48E-06	1 697	1 323	2 177	*ZNF236*	−47836
20	rs6077251*	7752366	T	C	0 271	0 153	2 50E-06	3 64E-06	3 06E-02	2 50E-06	2 065	1 537	2 773	*SFRS13AP2*	59982
8	rs11774576*	27740417	A	G	0 702	0 581	6 78E-05	2 82E-06	2 62E-01	2 82E-06	1 699	1 307	2 208	*SCARA5*	0
11	rs4627050*	18822037	G	A	0 781	0 649	4 43E-06	9 01E-06	2 17E-02	4 43E-06	1 932	1 452	2 572	*PTPN5*	−8648
1	rs12142335*	108302922	A	G	0 040	0 004	9 91E-06	8 45E-06	1 00E+00	8 45E-06	9 713	3 175	29 710	*VAV3*	0
Carbop at n															
15	rs11071200*	55950082	T	G	0 060	0 008	1 25E-06	8 51E-07	1 00E+00	8 51E-07	8 241	2 888	23 520	*PRTG*	0
5	rs3822735*	35799994	G	A	0 862	0 752	7 24E-06	1 68E-06	3 50E-01	1 68E-06	2 062	1 500	2 834	*SPEF2*	0
3	rs1623879*	58027197	G	A	0 441	0 321	7 69E-05	1 89E-02	3 75E-06	3 75E-06	1 669	1 297	2 148	*FLNB*	0
15	rs936229*	75132319	G	A	0 713	0 595	7 27E-05	4 41E-06	3 01E-01	4 41E-06	1 685	1 302	2 180	*ULK3*	0
13	rs7989332*	21050575	G	T	0 853	0 738	5 47E-06	1 16E-04	1 74E-03	5 47E-06	2 056	1 507	2 806	*CRYL1*	0
3	rs3845905*	66525963	G	A	0 921	0 828	5 99E-06	2 20E-05	1 50E-02	5 99E-06	2 433	1 645	3 598	*LRIG1*	0
8	rs1714746*	4105147	G	A	0 554	0 435	1 59E-04	6 63E-02	7 44E-06	7 44E-06	1 610	1 261	2 056	*CSMD1*	0
16	rs12446319*	81774798	A	G	0 253	0 143	8 97E-06	1 27E-04	3 16E-04	8 97E-06	2 026	1 480	2 774	*CMIP*	29431
1	rs1277203*	109392837	A	G	0 730	0 626	3 50E-04	3 51E-02	9 38E-06	9 38E-06	1 615	1 243	2 098	*AKNAD1*	0
A anthracyc ne-based drugs															
5	rs10040979*	158424391	G	A	0 701	0 618	4 68E-03	5 35E-01	4 60E-07	4 60E-07	1 452	1 120	1 883	*EBF1*	0
2	rs12615435*	200638509	T	G	0 883	0 773	3 95E-06	4 09E-06	9 17E-02	3 95E-06	2 214	1 555	3 154	*LOC348751*	0
5	rs7720283*	158459721	C	T	0 775	0 706	1 29E-02	3 37E-01	4 15E-06	4 15E-06	1 431	1 078	1 898	*EBF1*	0
1	rs1367448*	68633924	C	T	0 633	0 526	5 02E-04	5 32E-06	5 12E-01	5 32E-06	1 554	1 212	1 993	*LOC100289178*	0
6	rs2505059*	98495952	G	A	0 538	0 398	5 41E-06	2 29E-04	2 04E-04	5 41E-06	1 765	1 383	2 252	*MIR2113*	23455
12	rs4149639*	6442001	C	T	0 120	0 047	7 39E-06	1 52E-05	8 16E-02	7 39E-06	2 763	1 781	4 287	*TNFRSF1A*	0
19	rs1654260*	20329111	A	G	0 625	0 488	8 49E-06	3 03E-03	2 18E-05	8 49E-06	1 749	1 365	2 240	*LOC100421704*	−3576
Doxorub c n															
15	rs11857176*	78164706	A	G	0 657	0 515	1 74E-02	1 00E+00	8 08E-07	8 08E-07	1 800	1 127	2 874	*LOC100302666*	−6274
2	rs4380275*	773278	G	A	0 392	0 152	4 99E-06	1 54E-05	2 05E-02	4 99E-06	3 604	2 041	6 365	*LOC339822*	6559
11	rs2512987*	86414282	T	C	0 681	0 417	7 02E-06	6 04E-03	3 42E-05	7 02E-06	2 985	1 855	4 803	*ME3*	−30604
Ep rub c n															
12	rs4149639*	6442001	C	T	0 163	0 042	2 89E-07	8 31E-07	3 32E-02	2 89E-07	4 443	2 571	7 677	*TNFRSF1A*	0
5	rs2964475*	5407814	C	A	0 615	0 415	4 13E-06	3 95E-04	1 10E-04	4 13E-06	2 248	1 592	3 174	*KIAA0947*	−14993
13	rs1923834*	28360487	G	A	0 916	0 770	9 40E-06	4 61E-06	3 33E-01	4 61E-06	3 236	1 823	5 744	*GSX1*	−6293
10	rs908366*	126144839	A	G	0 518	0 328	7 04E-06	1 08E-04	8 69E-04	7 04E-06	2 199	1 564	3 092	*LHPP*	−5502

© 2013 Japanese Cancer Association
doi: 10.1111/cas.12186

Table 2. (continued)

CHR	SNP	BP	RA	NRA	RAF_Case	RAF_Ctr	P_a e c	P_dom	P_rec	Pm n	OR	L95	U95	*Gene*	re oc
3	rs1553091*	187716886	G	A	0 452	0 358	2 65E-02	8 04E-01	7 46E-06	7 46E-06	1 480	1 053	2 080	*LOC100505844*	−22691
A ant metabo te drugs															
18	rs7228133*	4539085	C	A	0 733	0 686	2 26E-02	6 64E-01	1 70E-06	1 70E-06	1 255	1 034	1 522	*LOC284215*	243085
21	rs8127977*	26826514	A	G	0 804	0 722	1 30E-05	2 11E-06	2 67E-01	2 11E-06	1 587	1 282	1 966	*NCRNA00158*	−22501
12	rs894734*	54319727	G	A	0 849	0 776	3 84E-05	3 97E-06	6 63E-01	3 97E-06	1 619	1 279	2 050	*HOXC13*	−12849
13	rs9580312*	22754093	G	A	0 480	0 409	1 35E-03	8 09E-06	6 25E-01	8 09E-06	1 330	1 120	1 581	*LOC100506622*	−30331
21	rs2055011*	19481354	C	T	0 184	0 143	1 12E-02	2 12E-01	8 82E-06	8 82E-06	1 347	1 075	1 686	*CHODL*	−135796
12	rs12582168*	124894184	C	T	0 333	0 256	8 50E-05	9 31E-06	2 84E-01	9 31E-06	1 454	1 210	1 748	*NCOR2*	0
5-F uorourac															
7	rs10488226*	12713070	A	C	0 195	0 107	1 09E-05	3 54E-06	2 98E-01	3 54E-06	2 026	1 500	2 737	*LOC100505995*	−12175
2	rs6740660*	224943685	G	A	0 966	0 894	4 10E-06	8 83E-06	2 40E-01	4 10E-06	3 386	1 870	6 131	*SERPINE2*	−39649
4	rs1567482*	36026747	G	A	0 952	0 875	6 26E-06	1 44E-05	9 14E-02	6 26E-06	2 846	1 716	4 719	*LOC651644*	39948
2	rs6706693*	192465598	A	G	0 328	0 219	1 62E-05	2 12E-03	9 45E-06	9 45E-06	1 743	1 362	2 232	*OBFC2A*	−77200
Gemc tab ne															
18	rs9961113*	75605399	C	T	0 625	0 403	1 43E-06	3 83E-04	3 73E-05	1 43E-06	2 473	1 706	3 584	*LOC100421527*	−260017
5	rs2547917*	58713680	A	G	0 350	0 212	9 06E-04	8 79E-02	3 33E-06	3 33E-06	1 997	1 345	2 965	*PDE4D*	0
15	rs12900463*	85415386	C	T	0 219	0 115	2 24E-03	1 02E-01	4 03E-06	4 03E-06	2 154	1 342	3 457	*ALPK3*	0
22	rs9609078*	31153276	T	C	0 089	0 009	4 32E-06	9 97E-06	2 59E-01	4 32E-06	10 890	3 528	33 610	*OSBP2*	0
5	rs6863418*	173625154	A	G	0 175	0 055	1 37E-05	6 98E-06	4 55E-01	6 98E-06	3 623	2 042	6 429	*HMP19*	88972
20	rs6037430*	344079	G	A	0 894	0 730	9 74E-06	1 75E-05	7 92E-02	9 74E-06	3 109	1 805	5 359	*NRSN2*	8567
A ant m crotubu e drugs															
17	rs11651483*	12777402	C	T	0 729	0 665	1 69E-03	2 60E-01	3 37E-07	3 37E-07	1 357	1 120	1 643	*RICH2*	0
6	rs4235898*	77266188	A	G	0 830	0 740	1 05E-06	3 34E-06	6 50E-03	1 05E-06	1 718	1 377	2 142	*LOC100131680*	−103976
13	rs4771859*	93088651	G	A	0 764	0 709	5 49E-03	2 60E-01	1 47E-06	1 47E-06	1 328	1 088	1 623	*GPC5*	0
1	rs12145418*	216716320	T	G	0 334	0 274	3 04E-03	3 17E-01	2 35E-06	2 35E-06	1 331	1 104	1 604	*ESRRG*	0
6	rs9386485*	106329055	T	C	0 596	0 492	2 65E-06	2 85E-05	8 59E-04	2 65E-06	1 524	1 279	1 817	*PRDM1*	−205140
16	rs12935229*	77328895	A	G	0 249	0 182	1 83E-04	2 23E-02	4 40E-06	4 40E-06	1 495	1 214	1 840	*ADAMTS18*	0
7	rs6961860*	17085321	G	A	0 557	0 495	5 33E-03	8 87E-01	4 67E-06	4 67E-06	1 283	1 078	1 527	*LOC100131425*	156806
14	rs12882718*	86902054	T	C	0 737	0 643	5 91E-06	5 55E-06	2 03E-02	5 55E-06	1 555	1 283	1 884	*LOC100421119*	−42891
12	rs1043763*	122630909	T	C	0 668	0 574	1 36E-05	6 51E-06	3 10E-02	6 51E-06	1 496	1 248	1 795	*MLXIP*	1920
2	rs4591358*	196365890	C	T	0 302	0 215	6 60E-06	6 89E-04	1 89E-04	6 60E-06	1 578	1 297	1 920	*LOC391470*	81627
14	rs8022296*	97987857	G	A	0 663	0 608	1 06E-02	6 54E-01	7 29E-06	7 29E-06	1 269	1 058	1 521	*LOC100129345*	111127
4	rs6817170*	154374984	G	A	0 377	0 286	9 29E-06	4 61E-05	3 27E-03	9 29E-06	1 517	1 264	1 822	*KIAA0922*	−12514
Pac taxe															
1	rs922106*	90025519	T	G	0 298	0 202	2 17E-04	1 95E-02	9 28E-07	9 28E-07	1 679	1 277	2 207	*LRRC8B*	0
6	rs9386485*	106329055	T	C	0 624	0 477	1 17E-06	1 43E-05	5 24E-04	1 17E-06	1 821	1 429	2 320	*PRDM1*	−205140
8	rs2444896*	99022009	T	G	0 727	0 603	1 58E-05	2 43E-06	7 41E-02	2 43E-06	1 754	1 355	2 269	*MATN2*	0
2	rs4666360*	20335709	C	T	0 216	0 114	4 62E-06	3 24E-06	2 27E-01	3 24E-06	2 136	1 546	2 950	*RPS16P2*	19625
9	rs3138083*	35648950	A	G	0 220	0 117	3 78E-06	1 24E-05	1 09E-02	3 78E-06	2 136	1 551	2 942	*SIT1*	345
17	rs3786094*	9875205	C	T	0 528	0 422	5 30E-04	1 80E-01	5 83E-06	5 83E-06	1 531	1 206	1 944	*GAS7*	0
15	rs4886670*	75449674	A	C	0 353	0 229	7 26E-06	5 38E-05	4 26E-03	7 26E-06	1 835	1 412	2 383	*RPL36AP45*	29318
5	rs792975*	172271007	T	C	0 654	0 519	7 66E-06	1 19E-04	6 08E-04	7 66E-06	1 746	1 366	2 232	*ERGIC1*	0
Docetaxe															
9	rs3747851*	124521260	T	C	0 337	0 176	5 61E-07	1 12E-05	5 63E-04	5 61E-07	2 377	1 693	3 339	*DAB2IP*	0
7	rs4727963*	122759980	C	T	0 772	0 618	7 99E-06	1 04E-06	1 69E-01	1 04E-06	2 094	1 505	2 914	*SLC13A1*	0
14	rs1756650*	87741025	G	A	0 211	0 162	9 95E-02	9 10E-01	1 74E-06	1 74E-06	1 386	0 954	2 014	*GALC*	658333

© 2013 Japanese Cancer Association

Table 2. (continued)

CHR	SNP	BP	RA	NRA	RAF_Case	RAF_Ctr	P_a e c	P_dom	P_rec	Pm n	OR	L95	U95	*Gene*	re oc
13	rs488248*	106596719	T	C	0 918	0 795	3 29E-06	3 23E-05	1 17E-02	3 29E-06	2 896	1 802	4 655	*LOC728192*	−432192
6	rs12660691*	130008445	A	C	0 935	0 819	3 62E-06	4 99E-06	1 62E-01	3 62E-06	3 199	1 899	5 391	*ARHGAP18*	0
18	rs4553720*	62170726	T	C	0 377	0 281	7 77E-03	4 56E-01	6 77E-06	6 77E-06	1 547	1 131	2 116	*LOC284294*	79890
6	rs2157460*	130021128	T	C	0 932	0 820	7 07E-06	1 07E-05	1 62E-01	7 07E-06	3 013	1 806	5 025	*ARHGAP18*	0
2	rs837841*	130034012	T	G	0 714	0 662	1 49E-01	5 26E-01	8 34E-06	8 34E-06	1 279	0 930	1 759	*LOC151121*	−33653
A topo somerase nh b tors															
5	rs10074959*	104208013	T	C	0 321	0 237	3 23E-02	8 08E-01	1 13E-06	1 13E-06	1 524	1 048	2 217	*RAB9P1*	−227162
7	rs1035147*	12094966	T	G	0 981	0 877	4 01E-06	4 75E-06	5 56E-01	4 01E-06	7 294	2 587	20 559	*TMEM106B*	−155882
1	rs303386*	99589379	A	G	0 585	0 444	1 10E-03	4 30E-06	3 88E-01	4 30E-06	1 766	1 256	2 483	*LOC100129620*	0
14	rs7494275*	56231800	C	A	0 543	0 406	1 86E-03	4 26E-01	8 50E-06	8 50E-06	1 732	1 232	2 433	*RPL13AP3*	−1163
3	rs480409*	7010081	G	A	0 495	0 348	6 08E-04	1 63E-01	9 28E-06	9 28E-06	1 842	1 307	2 596	*GRM7*	0
Camptothec n															
6	rs17318866*	3837198	G	A	0 966	0 790	1 47E-06	1 78E-06	1 91E-01	1 47E-06	7 559	2 689	21 263	*FAM50B*	−12434
2	rs17027130*	41273631	C	T	0 644	0 387	2 54E-06	8 03E-04	3 91E-05	2 54E-06	2 865	1 844	4 452	*LOC729984*	−110074
1	rs303386*	99589379	A	G	0 627	0 429	3 34E-04	3 61E-06	1 06E-01	3 61E-06	2 238	1 448	3 460	*LOC100129620*	0
Etopos de															
20	rs6039763*	10183517	A	G	0 370	0 090	1 27E-05	1 54E-06	4 61E-01	1 54E-06	5 966	2 502	14 230	*LOC100131208*	0
1	rs2506991*	48098406	G	A	0 593	0 269	1 39E-05	1 11E-02	2 28E-06	2 28E-06	3 948	2 101	7 418	*LOC388630*	127794
2	rs12987465*	49715021	A	G	0 593	0 359	1 87E-03	6 26E-01	6 61E-06	6 61E-06	2 597	1 424	4 737	*FSHR*	−333355
7	rs3095008*	20255705	T	C	1 000	0 846	1 74E-05	9 39E-06	1 00E+00	9 39E-06	nf	N/A	N/A	*MACC1*	0
Pac taxe + carbop at n															
12	rs12310399*	95490248	A	G	0 708	0 567	2 60E-04	1 08E-01	2 46E-07	2 46E-07	1 852	1 329	2 580	*FGD6*	0
9	rs10785877*	137125501	T	C	0 833	0 660	7 38E-07	1 54E-05	1 98E-04	7 38E-07	2 580	1 766	3 769	*RXRA*	-92815
19	rs995834*	28866596	C	T	0 580	0 425	1 28E-04	1 20E-06	1 51E-01	1 20E-06	1 871	1 364	2 566	*LOC100420587*	307385
1	rs922107*	90022796	G	A	0 323	0 211	1 55E-03	9 07E-02	2 73E-06	2 73E-06	1 788	1 250	2 558	*LRRC8B*	0
7	rs1425132*	37562368	T	C	0 740	0 666	4 55E-02	7 35E-01	4 68E-06	4 68E-06	1 430	1 013	2 017	*LOC442668*	62349
1	rs6429703*	15339960	T	C	0 200	0 078	8 40E-06	2 59E-05	5 61E-02	8 40E-06	2 942	1 802	4 804	*RP1-21O18.1*	0

*SNPs used for we ghted genet c r sk score ana yses BP, SNP genom c ocat on; CHR, chromosome; nf, nfin ty; L95, ower 95% confidence nterva ; N/A, not app cab e; NRA, non-r sk a e e; OR, odds rat o; P_a e c, *P*-va ue from a e c mode ; P_dom, *P*-va ue from dom nant mode ; P_m n, m n mum *P*-va ue among the three mode s; P_rec, *P*-va ue from recess ve mode ; RA, r sk a e e; RAF, r sk a e e frequency; re oc , d stance of the SNP from the gene; U95, upper 95% confidence nterva

doi: 10.1111/cas.12186

© 2013 Japanese Cancer Association

Among these datasets, GWAS carried out using samples who were given: (i) any kind of platinum based chemotherapy (428 cases *vs* 743 controls); (ii) cisplatin based chemotherapy (176 cases *vs* 471 controls); or (iii) carboplatin based chemo therapy (261 cases *vs* 262 controls) identified SNPs showing the most significant association with chemotherapy induced severe neutropenia/leucopenia are: rs4886670 (P_{min} = 9.86 × 10^{-7}, OR = 1.61, 95% CI = 1.33 1.94) near *RPL36AP45* for (i); rs10253216 (P_{min} = 1.68 × 10^{-7}, OR = 1.48, 95% CI = 1.16 1.89) near *AGR2* for (ii); and rs11071200 (P_{min} = 8.51 × 10^{-7}, OR = 8.24, 95% CI = 2.89 23.5) on *PRTG* for (iii) (Table 2, Table S1, Fig. S2b). For the anthracycline based reg imen, we carried out GWAS with individuals given all anthra cycline based (184 cases *vs* 459 controls), doxorubicin based (83 cases *vs* 66 controls), and epirubicin based (83 cases *vs* 370 controls) chemotherapy, and identified three SNPs, rs10040979 (P_{min} = 4.60 × 10^{-7}, OR = 1.45, 95% CI = 1.12 1.88) in *EBF1*, rs11857176 (P_{min} = 8.08 × 10^{-7}, OR = 1.80, 95% CI = 1.13 2.87) near a hypothetical gene *LOC10030 2666*, and rs4149639 (P_{min} = 2.89 × 10^{-7}, OR = 4.44, 95% CI = 2.57 7.68) in *TNFRSF1A*, to be most significantly associ ated with the risk of high grade neutropenia/leucopenia, respectively (Table 2, Table S1, Fig. S2c). In the case of an timicrotubule agents, we carried out three different GWAS with individuals who were treated with antimicrotubule (371 cases *vs* 825 controls), paclitaxel based (218 cases *vs* 364 con trols), or docetaxel based (147 cases *vs* 233 controls) regimens. We identified three SNPs, rs11651483 (P_{min} = 3.37 × 10^{-7}, OR = 1.36, 95% CI = 1.12 1.64) in *RICH2*, rs922106 (P_{min} = 9.28 × 10^{-7}, OR = 1.68, 95% CI = 1.28 2.21) in *LRRC8B* and rs3747851 (P_{min} = 5.61 × 10^{-7}, OR = 2.38, 95% CI = 1.69 3.34) in *DAB2IP*, to be those most signifi cantly associated with the increased risk of severe neutropenia /leucopenia, respectively (Table 2, Table S1, Fig. S2e). Our previous report by Kiyotani *et al.*[24] identified four SNPs to be associated with gemcitabine induced hematological toxici ties. Three of the four SNPs were included in the current study with suggestive association, rs12046844 (P_{min} = 5.84 × 10^{-4}, OR = 2.53, 95% CI = 1.45 4.43), rs6430443 (P_{min} = 8.61 × 10^{-4}, OR = 6.33, 95% CI = 1.90 22.2; r^2 = 0.895 with rs1901440) and rs11719165 (P_{min} = 1.16 × 10^{-2}, OR = 2.36, 95% CI = 1.18 4.70) (Table S4). However, it is noted that some of the samples used in this study overlapped with those in the study reported by Kiyotani *et al.*, as both sourced sam ples from Biobank Japan.

Lastly, we also attempted to identify genetic variants associ ated with combined treatment of paclitaxel and carboplatin induced severe neutropenia/leucopenia (150 cases *vs* 166 controls), as this combined treatment is commonly used as the standard therapy for both ovarian and lung cancers. We found the most significant association with the SNP rs12310399 (P_{min} = 2.46 × 10^{-7}, OR = 1.85, 95% CI = 1.33 2.58) near the *FGD6* gene (Table 2, Table S1, Fig. S2a), which is suggested to activate CDC42, a member of the Ras like family of Rho and Rac proteins, and has a critical role in regulating the actin cytoskeleton. The second strongest association was observed at the locus encoding RXRA (P_{min} = 7.38 × 10^{-7}, OR = 2.58, 95% CI = 1.77 3.77), an important transcriptional factor. We also calculated the cumulative genetic scores using SNPs on six loci and identified that individuals in group 4 could have 188 times (95% CI = 36.1 979) higher risk of developing severe neutropenia/leucopenia than those belonging to group 1 with the sensitivity of 95.9% and the specificity of 88.9% (Table S2). Because this drug combination is of clinical importance, we further investigated the association of these six selected loci using 161 individuals who developed grade 1/2 neutropenia/leucopenia, using cases registered in the Biobank Japan. Interestingly, the association results for the six loci were moderate for grade 1/2 neutropenia/leucopenia, with intermediate allele frequency and OR between individuals without any adverse reactions and those with neutropenia/leu copenia of ≥grade 3 (Table S3). In addition, as shown in Table 3 and Figure 1, the higher the calculated score becomes, the higher the proportion and grade of neutropenia/leucopenia. The intermediate scores for patients with grade 1/2 neutrope nia/leucopenia could imply the possible usefulness of this scoring system for the prediction.

Furthermore, we used simulation to estimate how many samples are required to validate this scoring result. We started off by estimating the incidence of neutropenia/leucopenia by the combined treatment of paclitaxel and carboplatin. In Bio bank Japan, a total of 477 individuals received this combined treatment; among them, 166 individuals (35%) did not develop any adverse drug reactions, 161 (35%) developed mild neutro penia/leucopenia (grade 1 or 2) and 150 (30%) developed severe neutropenia/leucopenia (grade 3 or higher). The frequency of developing severe neutropenia/leucopenia is in agreement with a multicenter study reported by Guastalla *et al.*[25] When we assume that 100 patients who receive this combination therapy are prospectively registered, the inci dences of the adverse drug reactions are estimated as shown in Table 4. If we categorize the patients by wGRS according to the proportions indicated in Table 3 (and our hypothesis is right), the statistical power should be enough to validate by this small subset of patients. Even if two individuals in both group 1 and group 4 are incorrectly predicted, the calculated *P* value is still 0.03 by Fisher's exact test.

Discussion

In this study, we carried out GWAS analyses for a total of 17 subsets of chemotherapies to identify genetic variants that might be associated with chemotherapeutic induced neutrope nia/leucopenia with grades 3 and 4, however, we could not identify any SNPs that surpassed the genome wide significant threshold (*P* value < 5 × 10^{-8}). Through this study, we

Table 3. Weighted genetic risk score (wGRS) analysis of cancer patients who received combination treatment with paclitaxel and carboplatin

wGRS group	Score	G3G4	G1G2	G0	% G3G4	% G1G2	% G0	G3/4 versus G0			G1/2 versus G0		
								OR	95% CI	*P* value	OR	95% CI	*P* value
1	<5.802	2	21	48	0.03	0.29	0.68		REF			REF	
2	5.802 7.665	36	58	77	0.21	0.34	0.45	11.2	2.58 48.70	8.69E 05	1.72	0.93 3.19	9.55E 02
3	7.665 9.528	64	62	33	0.40	0.39	0.21	46.5	10.60 204.00	2.36E 13	4.29	2.21 8.35	1.61E 05
4	>9.528	47	20	6	0.64	0.28	0.08	188.0	36.10 979.00	4.78E 20	7.62	2.68 21.70	6.08E 05
Total		149	161	164									

95% CI, 95% confidence interval; G0, individuals who did not develop any adverse drug reaction; G1G2, grade 1 and grade 2 neutropenia (mild); G3G4, grade 3 and grade 4 neutropenia (severe); OR, odds ratio; REF, reference.

© 2013 Japanese Cancer Association

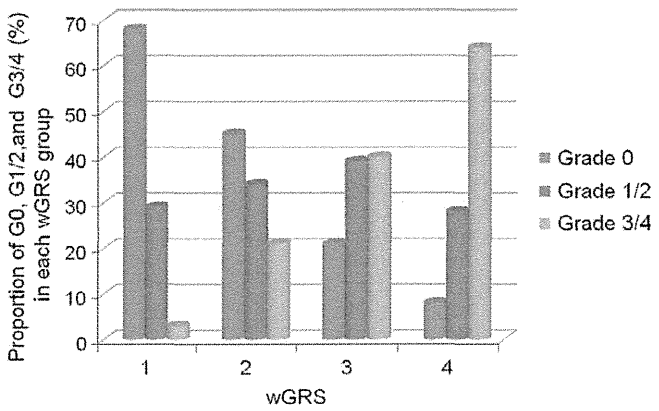

Fig. 1. Proportions of cancer patients who developed no adverse reaction (G0), mild neutropenia/leucopenia (G1/2), or severe neutropenia/leucopenia (G3/4) in each of the weighted genetic risk score (wGRS) score groups. All patients received combined treatment with paclitaxel and carboplatin and were registered with Biobank Japan. The total numbers of patients in scores 1, 2, 3, and 4 are 71, 171, 159, and 73, respectively.

Table 4. Simulation of weighted genetic risk score (wGRS) analysis for a prospective study of 100 patients who received combination treatment with paclitaxel and carboplatin

Estimated verification samples (*n* 100; 35 expected to have grade 1/2 neutropenia)					
wGRS group	G3G4	G0	OR	95% CI	*P* value
1	0	10			
2	7	17	9.0	0.47 174.00	7.82E 02
3	13	7	37.8	1.93 740.00	1.06E 03
4	10	1	147.0	5.35 4040.00	3.40E 05
Total	30	35			

95% CI, 95% confidence interval; G0, individuals without any adverse drug reaction; G3G4, grade 3 and 4 neutropenia (severe); OR, odds ratio.

encountered several important issues, which are now common problems in pharmacogenomics studies using retrospective clinical data, including confounding factors and heterogeneous treatments for individual patients (often given different combi nations of drugs, different dosage of drugs, and different time periods of treatment), that increase the complexity of studies and generate various noises in the analyses, and diminished the statistical power in the case control association studies. We understand that our current approach was not an ideal study design, but it is not easy to perfectly standardize therapy in the daily clinical practice of cancer treatment. There are several factors contributing to the variability in treatments: (i) there is some preference by doctors or by hospitals to select a particular regimen among the various recommended standard treatments; (ii) the modifications (adjustments) of the dosage or schedule according to the patient's conditions (performance status, results of laboratory tests, etc.); and (iii) although we have been collecting the clinical information, it is not perfect to collect complete clinical information in some hospitals, particularly those that do not use electronic medical records. One can say that this kind of study should be performed as a prospective design, however, due to the very rapid advances in the development of novel molecular targeted drugs and new regimens in the oncology area, the protocols have been and will be modified or improved. Hence, spending many years and a huge budget on a prospective study may result in a clini cally useless outcome, because the results are unable to be applied due to the replacement of the study protocol with a new protocol, when the results of association studies are avail able. Nevertheless, retrospective pharmacogenomic studies could be improved by implementing electronic medical record systems that could include detailed descriptions of patients' conditions and their responses to various drugs.

Although we understand the pitfalls in study designs like our present study, we need to seek possible ways to identify candi date genetic variants that might contribute to improvement in the clinical management of cancer patients, including chemo therapy induced severe neutropenia/leucopenia. Nevertheless, some of the candidate genes that we identified are of interest, considering their known functions as well as their relations with drug actions. For example, the proto oncogene *AGR2*, whose genetic variants were suggested to associate with cis platin induced neutropenia/leucopenia, encodes an anterior gradient 2 homolog (*Xenopus laevis*) that is known to play a critical role in cell migration, cell differentiation, and cell growth.[26] Cells stably expressing *AGR2* confer resistance to cisplatin *in vivo*, compared with control cells (empty vector) in a xenograft animal model.[27] The second example is *TNFRSF1A*, suggested to be associated with anthracycline based and epirubicin induced neutropenia/leucopenia. This gene encodes TNFRSF1A, which is a major receptor for TNF α. The soluble TNFRSF1A level was found to be elevated after 1 month of anthracycline based chemotherapy.[28] Addi tionally, both TNF α and TNFRSF1A are known to play a crit ical role in doxorubicin induced cardiotoxicity, in which doxorubicin stimulates an increase in circulating TNF and upregulates TNFRSF1A.[29,30] Furthermore, genetic variants on *PDE4D*, which encodes for phosphodiesterase 4D, cAMP specific, showed suggestive association with gemcitabine induced severe neutropenia/leucopenia. Ablation of *PDE4D* has been reported to impair the neutrophil function with altered chemotaxis ability and adhesion capability as well as to reduce neutrophil recruitment to the site of inflammation.[31] Besides, genetic variants on *RXRA* identified to be associated with combined treatment of paclitaxel and carboplatin induced severe neutropenia/leucopenia, encodes retinoid X receptor alpha. Disruption of this gene in mouse models moderately alters lymphocyte proliferation and survival, and affects the T helper type1/type 2 balances.[32] All of these genes might pro vide some important insights into the mechanism of various chemotherapy induced severe neutropenia/leucopenia, how ever, further validations are definitely essential.

As already described, the GWAS approach could provide a list of genetic variants that might be associated with complex phenotypes (drug responsiveness or drug induced adverse reactions) in pharmacogenomics studies. One of the clinically important aims for identification of the associated genetic variants is to establish a prediction model to identify individ uals who are at risk of adverse reactions with certain drugs or protocols. In this study, we have applied the wGRS sys tem, by which we could distinguish high risk patients from low risk individuals by counting the number of risk alleles of the suggestively associated SNPs in combination with esti mating the effect size of each SNP. One of the best examples from this study was indicated by a scoring system using six candidate SNP loci that were identified through the GWAS of severe neutropenia/leucopenia caused by combination treatment of paclitaxel and carboplatin; among 53 individuals in the high risk group (group 4) by this scoring method, 47 (89%) revealed high grade neutropenia/leucopenia. In con trast, among 50 individuals in the low risk group (group 1), only 2 (4%) revealed high grade neutropenia/leucopenia, and

doi: 10.1111/cas.12186
© 2013 Japanese Cancer Association

the odds ratio to have the severe adverse reaction in individuals belonging to group 4 was calculated to be 188 times higher than those categorized to group 1 (Table 3). Interestingly, individuals who developed grade 1/2 (mild neutropenia/leucopenia) were found to show intermediate risk scores between patients with severe neutropenia/leucopenia and those without any adverse reactions. Hence, we suggest that wGRS is an applicable method to evaluate the clinical utility of possible variants with specific phenotypes. However, the data are preliminary and require verification by an inde pendent test sample(s) before any definitive conclusions can be drawn. But, considering that the OR of the high risk group is very high, the number of samples required for the verification (if our hypothesis is right) is not so large. In fact, we have tried to simulate a prospective study design using a model of 100 patients according to the assumption that 35% individuals will not develop any adverse drug reactions, 35% individuals will develop mild neutropenia/leucopenia (grade 1/2), and 30% will develop severe neutropenia/leucopenia (grade 3/4). As shown in Table 4, the study of 100 patients should have very strong statistical power to verify. If this is verified, as we expect, it should improve the quality of lives of cancer patients and also contribute to reducing medical care costs by avoiding unnecessary adverse events. However, to achieve success in pharmacogenomics and personalized medicine, both local and international collaborative efforts are essential.

Acknowledgments

We would like to express our heartfelt gratitude to all the patients who participated in this study. We convey our sincere appreciation to Dr Teruhiko Yoshida and Dr Hiromi Sakamoto from the National Can cer Center Research Institute for their kind support. Our thanks also goes to the members of the laboratory for statistical analysis and the laboratory for the genotyping development from the Center for Geno mic Medicine for their kind support and fruitful discussions. We would like to extend our gratitude to the staff of Biobank Japan for their out standing assistance. This work was carried out as part of the Biobank Japan Project, supported by the Ministry of Education, Culture, Sports, Sciences and Technology, Japan. In addition, this project was supported by the JSPS postdoctoral fellowship.

Disclosure Statement

The authors have no conflict of interest.

Abbreviations

GWAS	genome wide association study
OR	odds ratio
SD	standard deviation
SNP	single nucleotide polymorphism
TNFRSF1A	tumor necrosis factor receptor superfamily, member 1A
TNF α	tumor necrosis factor α
wGRS	weighted genetic risk score

References

1 Wheeler HE, Maitland ML, Dolan ME, Cox NJ, Ratain MJ. Cancer pharmacogenomics: strategies and challenges. *Nat Rev Genet* 2013; **14**: 23 34.
2 Relling MV, Hancock ML, Rivera GK *et al.* Mercaptopurine therapy intolerance and heterozygosity at the thiopurine S-methyltransferase gene locus. *J Natl Cancer Inst* 1999; **91**: 2001 8.
3 Weinshilboum RM, Sladek SL. Mercaptopurine pharmacogenetics: monogenic inheritance of erythrocyte thiopurine methyltransferase activity. *Am J Hum Genet* 1980; **32**: 651 62.
4 Relling MV, Klein TE. CPIC: clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin Pharmacol Ther* 2011; **89**: 464 7.
5 Innocenti F, Undevia SD, Iyer L *et al.* Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan. *J Clin Oncol* 2004; **22**: 1382 8.
6 Iyer L, Das S, Janisch L *et al.* UGT1A1*28 polymorphism as a determinant of irinotecan disposition and toxicity. *Pharmacogenomics J* 2002; **2**: 43 7.
7 Hoskins JM, Goldberg RM, Qu P, Ibrahim JG, McLeod HL. UGT1A1*28 genotype and irinotecan-induced neutropenia: dose matters. *J Natl Cancer Inst* 2007; **99**: 1290 5.
8 McCarthy MI, Abecasis GR, Cardon LR *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008; **9**: 356 69.
9 Triendl R. Japan launches controversial Biobank project. *Nat Med* 2003; **9**: 982.
10 Takata R, Akamatsu S, Kubo M *et al.* Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat Genet* 2010; **42**: 751 4.
11 Akamatsu S, Takata R, Haiman CA *et al.* Common variants at 11q12, 10q26 and 3p11.2 are associated with prostate cancer susceptibility in Japanese. *Nat Genet* 2012; **44**: 426 9, S1.
12 Cui R, Kamatani Y, Takahashi A *et al.* Functional variants in ADH1B and ALDH2 coupled with alcohol and smoking synergistically enhance esophageal cancer risk. *Gastroenterology* 2009; **137**: 1768 75.
13 Cui R, Okada Y, Jang SG *et al.* Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population. *Gut* 2011; **60**: 799 805.
14 Kumar V, Kato N, Urabe Y *et al.* Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. *Nat Genet* 2011; **43**: 455 8.
15 Low SK, Kuchiba A, Zembutsu H *et al.* Genome-wide association study of pancreatic cancer in Japanese population. *PLoS ONE* 2010; **5**: e11824.
16 Elgazzar S, Zembutsu H, Takahashi A *et al.* A genome-wide association study identifies a genetic variant in the SIAH2 locus associated with hormonal receptor-positive breast cancer in Japanese. *J Hum Genet* 2012; **57**: 766 71.
17 Miki D, Ochi H, Hayes CN *et al.* Variation in the DEPDC5 locus is associated with progression to hepatocellular carcinoma in chronic hepatitis C virus carriers. *Nat Genet* 2011; **43**: 797 800.
18 Miki D, Kubo M, Takahashi A *et al.* Variation in TP63 is associated with lung adenocarcinoma susceptibility in Japanese and Korean populations. *Nat Genet* 2010; **42**: 893 6.
19 Cha PC, Zembutsu H, Takahashi A, Kubo M, Kamatani N, Nakamura Y. A genome-wide association study identifies SNP in DCC is associated with gallbladder cancer in the Japanese population. *J Hum Genet* 2012; **57**: 235 7.
20 Silvestris N, Del Re M, Azzariti A *et al.* Optimized granulocyte colony-stimulating factor prophylaxis in adult cancer patients: from biological principles to clinical guidelines. *Expert Opin Ther Targets* 2012; **16**(Suppl 2): S111 7.
21 Yamaguchi-Kabata Y, Nakazono K, Takahashi A *et al.* Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am J Hum Genet* 2008; **83**: 445 56.
22 Low SK, Takahashi A, Cha PC *et al.* Genome-wide association study for intracranial aneurysm in the Japanese population identifies three candidate susceptible loci and a functional genetic variant at EDNRA. *Hum Mol Genet* 2012; **21**: 2102 10.
23 De Jager PL, Chibnik LB, Cui J *et al.* Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *Lancet Neurol* 2009; **8**: 1111 9.
24 Kiyotani K, Uno S, Mushiroda T *et al.* A genome-wide association study identifies four genetic markers for hematological toxicities in cancer patients receiving gemcitabine therapy. *Pharmacogenet Genomics* 2012; **22**: 229 35.
25 Guastalla JP, Pujade-Lauraine E, Weber B *et al.* Efficacy and safety of the paclitaxel and carboplatin combination in patients with previously treated advanced ovarian carcinoma. A multicenter GINECO (Group d'Investigateurs Nationaux pour l'Etude des Cancers Ovariens) phase II study. *Ann Oncol* 1998; **9**: 37 43.
26 Wang Z, Hao Y, Lowe AW. The adenocarcinoma-associated antigen, AGR2, promotes tumor growth, cell migration, and cellular transformation. *Cancer Res* 2008; **68**: 492 7.
27 Hengel SM, Murray E, Langdon S *et al.* Data-independent proteomic screen identifies novel tamoxifen agonist that mediates drug resistance. *J Proteome Res* 2011; **10**: 4567 78.
28 Perik PJ, De Vries EG, Boomsma F *et al.* The relation between soluble apoptotic proteins and subclinical cardiotoxicity in adjuvant-treated breast cancer patients. *Anticancer Res* 2006; **26**: 3803 11.

© 2013 Japanese Cancer Association

29 Mukherjee S, Banerjee SK, Maulik M, Dinda AK, Talwar KK, Maulik SK. Protection against acute adriamycin-induced cardiotoxicity by garlic: role of endogenous antioxidants and inhibition of TNF-alpha expression. *BMC Pharmacol* 2003; **3**: 16.
30 Gilliam LA, St Clair DK. Chemotherapy-induced weakness and fatigue in skeletal muscle: the role of oxidative stress. *Antioxid Redox Signal* 2011; **15**: 2543 63.
31 Ariga M, Neitzert B, Nakae S *et al.* Nonredundant function of phosphodiesterases 4D and 4B in neutrophil recruitment to the site of inflammation. *J Immunol* 2004; **173**: 7531 8.
32 Stephensen CB, Borowsky AD, Lloyd KC. Disruption of Rxra gene in thymocytes and T lymphocytes modestly alters lymphocyte frequencies, proliferation, survival and T helper type 1/type 2 balance. *Immunology* 2007; **121**: 484 98.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Quantile quantile plots of 17 genome wide association studies of drug induced or drug subgroup induced severe neutropenia and their corresponding lambda (λ) value.

Fig. S2. (a) Manhattan plot for genome wide association study of cyclophosphamide and paclitaxel + carboplatin induced severe neutropenia/leuco penia. (b) Manhattan plot for genome wide association study of severe neutropenia/leucopenia induced by all types of platinum based agents, cisplatin, or carboplatin. (c) Manhattan plot for genome wide association study of severe neutropenia/leucopenia induced by all types of anthracy cline based agents, doxorubicin, or epirubicin. (d) Manhattan plot for genome wide association study of severe neutropenia/leucopenia induced by all types of antimetabolite agents, 5 fluorouracil, or gemcitabine. (e) Manhattan plot for genome wide association study of severe neutropenia induced by all types of antimicrotubule agents, paclitaxel, or docetaxel. (f) Manhattan plot for genome wide association study severe neutropenia/leucopenia induced by all types of topoisomerase inhibitors, camptothecin, or etoposide.

Table S1. Genome wide association study of each chemotherapy regimen with $P < 1 \times 10^{-4}$.

Table S2. Weighted genetic risk score of each genome wide association study of specific chemotherapeutic based induced severe neutropenia/leu copenia.

Table S3. Association study of cancer patients who do not develop any adverse drug reaction and those who developed neutropenia/leucopenia after being given combination treatment with paclitaxel and carboplatin.

Table S4. Association of previously reported SNPs that associated with gemcitabine induced hematological toxicity.

doi: 10.1111/cas.12186
© 2013 Japanese Cancer Association

Carcinogenesis vol.00 no.00 p.1 of 8, 2013
doi:10.1093/carcin/bgt374
Advance Access publication November 15, 2013

Association of microRNA-31 with *BRAF* mutation, colorectal cancer survival and serrated pathway

Katsuhiko Nosho[1,*,†], Hisayoshi Igarashi[1,†], Masanori Nojima[2,†], Miki Ito[1,†], Reo Maruyama[3], Shinji Yoshii[4,5], Takafumi Naito[1], Yasutaka Sukawa[1], Masashi Mikami[1], Wakana Sumioka[1], Eiichiro Yamamoto[1], Sei Kurokawa[1], Yasushi Adachi[1], Hiroaki Takahashi[1,4], Hiroyuki Okuda[1,4], Takaya Kusumi[6], Masao Hosokawa[6], Masahiro Fujita[7], Tadashi Hasegawa[8], Kenji Okita[9], Koichi Hirata[9], Hiromu Suzuki[3,†], Hiroyuki Yamamoto[10,†] and Yasuhisa Shinomura[1,†]

[1]Department of Gastroenterology, Rheumatology and Clinical Immunology, Sapporo Medical University School of Medicine, Sapporo, Japan, [2]Division of Advanced Medicine Promotion, The Advanced Clinical Research Center, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan, [3]Department of Molecular Biology, Sapporo Medical University School of Medicine, Sapporo, Japan, [4]Department of Gastroenterology, Keiyukai Sapporo Hospital, Sapporo, Japan, [5]Department of Gastroenterology, NTT East Sapporo Hospital, Sapporo, Japan, [6]Department of Surgery and [7]Department of Pathology, Keiyukai Sapporo Hospital, Sapporo, Japan, [8]Department of Clinical Pathology and [9]Department of Surgery, Surgical Oncology and Science, Sapporo Medical University School of Medicine, Sapporo, Japan and [10]Division of Gastroenterology and Hepatology, Department of Internal Medicine, St. Marianna University School of Medicine, Kawasaki, Japan

*To whom correspondence should be addressed. Department of Gastroenterology, Rheumatology and Clinical Immunology, Sapporo Medical University School of Medicine, S-1, W-16, Chou-ku, Sapporo 060-8543, Japan. Tel: +81(11)611-2111; Fax: +81(11)611-2282; Email: nosho@sapmed.ac.jp

BRAF* is an important gene in colorectal cancers (CRCs) that is associated with molecular characterization and resistance to targeted therapy. Although microRNAs (miRNAs) are useful biomarkers of various cancers, the association between miRNA and *BRAF* in CRCs is undefined. Therefore, this study was conducted to identify a relationship between specific miRNA molecules and *BRAF* mutation in CRCs and serrated lesions. miRNA array was used for the measurement of 760 miRNAs in 29 CRCs. To assess the identified miRNAs, quantitative reverse transcription–PCR was performed on 721 CRCs, 381 serrated lesions and 251 non-serrated adenomas. Moreover, proliferation and invasion assays were conducted using cell lines. miRNA array analysis revealed that microRNA-31 (miR-31)-5p was the most up-regulated miRNA in CRCs with mutated *BRAF* (*V600E*) compared with CRCs possessing wild-type *BRAF* (including cases with *KRAS* mutation). High miR-31 expression was associated with *BRAF* and *KRAS* mutations and proximal location ($P < 0.0001$). High miR-31 expression was related to cancer-specific mortality [multivariate hazard ratio = 2.06, 95% confidence interval: 1.36–3.09, $P = 0.0008$]. Functional analysis demonstrated that miR-31 inhibitor decreased cell invasion and proliferation. With regard to serrated lesions, high miR-31 expression was less frequently detected in hyperplastic polyps compared with other serrated lesions. In conclusion, associations were identified between miR-31, *BRAF* and prognosis in CRC. Transfection of miR-31 inhibitor had an antitumour effect. Thus, miR-31 may be a promising diagnostic biomarker and therapeutic target in colon cancers. Moreover, high miR-31 expression in serrated lesions suggested that miR-31 may be a key molecule in serrated pathway.

Abbreviations: CI, confidence interval; CRC, colorectal cancer; EGFR, epidermal growth factor receptor; FFPE, formalin-fixed paraffin-embedded; HP, hyperplastic polyp; HR, hazard ratio; miRNA, microRNA; miR-31, microRNA-31; MSI, microsatellite instability; MSS, microsatellite stability; OR, odds ratio; qRT–PCR, quantitative reverse transcription–PCR; SSA/P, sessile serrated adenoma/polyp; TSA, traditional serrated adenoma.

†These authors contributed equally to this work.

Introduction

BRAF, a member of the RAF gene family, which encodes a serine–threonine protein kinase and plays an important role in the activation of RAS–RAF–MEK–ERK signalling pathway, is one of the targeted genes in colorectal cancers (CRCs) (1–5). With regard to patient survival and chemoresistance, previous studies have shown an association between *BRAF* (*V600E*) mutation and high cancer-specific mortality rates among patients with CRC (2,5). In addition, *BRAF* has been associated with resistance to monoclonal antibodies against the epidermal growth factor receptor (EGFR) in patients with *KRAS* wild-type metastatic CRC refractory to chemotherapy (2,3). Therefore, these results suggest that further analysis of *BRAF* may enable us to identify the molecular characterization and the potential therapeutic target in CRCs.

MicroRNAs (miRNAs) constitute a class of small non-coding RNA molecules (21–25 nucleotides) that function as post-transcriptional gene regulators. miRNAs can function as oncogenes or tumour suppressors. Therefore, they have been increasingly recognized as useful biomarkers for various human cancers (6–15). In CRCs, several miRNAs are known to be deregulated (16–39) and target genes in the downstream effectors of EGFR (25,29,31–33). However, miRNAs specific to *BRAF* or its activation remain largely unknown.

The serrated pathway has attracted considerable attention as an alternative route to CRC. Approximately 30% of CRCs are hypothesized to arise from serrated lesions (40). Accumulating evidence suggests an association between CRCs with mutated *BRAF* and serrated lesions in many cases, as indicated by the high frequency of *BRAF* mutation in serrated lesions (40,41). Of these lesions, sessile serrated adenoma/polyp (SSA/P) has been identified as the precursor lesion of microsatellite instability (MSI)-high CRC with *BRAF* mutation in the proximal colon (40–42). However, the role of miRNAs in the development of CRC via the serrated pathway has not been examined in large samples of serrated lesions to date.

Therefore, we hypothesized that some specific miRNA molecules may regulate *BRAF* activation in CRCs. They may also play an important role in the progression of serrated lesions. To test this hypothesis, we conducted miRNA array analysis to detect miRNA molecules that are potentially associated with *BRAF* mutation using a database of 1353 colorectal tumours.

Materials and methods

Patients and tissue specimens

Formalin-fixed, paraffin-embedded (FFPE) tissues of 735 CRCs (stages I–IV), 391 serrated lesions and 259 non-serrated adenomas (i.e. tubular or tubulovillous adenomas) of patients who underwent endoscopic resection or other surgical treatment at Sapporo Medical University Hospital, Keiyukai Sapporo Hospital and JR Sapporo Hospital between 1997 and 2012 were collected. To avoid selection bias as far as possible, we consecutively collected FFPE specimens of CRC tissues, serrated lesions and non-serrated adenomas. The criterion for diagnosis of CRC was invasion of malignant cells beyond the muscularis mucosa. Intramucosal carcinoma and carcinoma *in situ* were classified as adenoma. Colorectal tumours were classified by location as follows: the proximal colon (caecum, ascending and transverse colon), distal colon (splenic flexure, descending and sigmoid colon) and rectum.

© The Author 2013. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

To clarify the association between miRNA expression and survival in metastatic CRC patients, we limited the patients who received adjuvant chemotherapy to those treated with 5-fluorouracil-based adjuvant chemotherapy. Patients with metastatic CRC treated with other targeted therapies (i.e. anti-vascular endothelial growth factor or anti-EGFR antibody) were excluded. The patients were followed up until death or December 2012, whichever came first. Informed consent was obtained from all the patients before specimen collection. This study was approved by the respective institutional review boards of the participating institutions. Tumour and paired normal colorectal tissues were reviewed by two pathologists (M.F. and T.H.). The term 'prognostic marker' is used throughout this article according to the REMARK Guidelines (43).

Histopathological evaluation of colorectal serrated lesions

Histological findings for all colorectal serrated lesion specimens were evaluated by a pathologist (M.F.) who was blinded to the clinical and molecular information. Serrated lesions [hyperplastic polyps (HPs) (N = 145), SSA/P (N = 131) and traditional serrated adenoma (TSA) (N = 115)] were classified on the basis of the current World Health Organization (WHO) criteria (44).

RNA extraction and miRNA array analysis

Total RNA was extracted from FFPE tissues using the miRNeasy FFPE Kit (Qiagen, Valencia, CA). The TaqMan® Array Human MicroRNA A + B Cards Set v3.0 (Applied Biosystems, Foster City, CA) was used for simultaneous measurement of the expression of 760 miRNAs on a microfluidic PCR platform. In brief, 1 μg of total RNA was reverse transcribed using the Megaplex Pools Kit (Applied Biosystems), following which miRNAs were amplified and detected by PCR with specific primers and TaqMan probes. PCR was run in the 7900HT Fast Real-Time PCR system (Applied Biosystems), and SDS 2.2.2 software (Applied Biosystems) was used for comparative analysis of the cycle threshold (ΔC_T). U6 snRNA (RNU6B; Applied Biosystems) served as an endogenous control. ΔC_T was calculated by subtracting the C_T values of U6 from the C_T values of the gene of interest. Expression of each miRNA in the tumour samples was calculated using the equation $2^{-\Delta C_T}$, where ΔC_T = ($C_T$miRNA – C_TU6).

Quantitative reverse transcription–PCR of miR-31

MicroRNA-31 (miR-31)-5p expression was analysed using TaqMan microRNA Assays (Applied Biosystems). In brief, 5 ng of total RNA were reverse transcribed using specific stem–loop RT primers, following which they were amplified and detected by quantitative reverse transcription–PCR (qRT–PCR) with specific primers and TaqMan probes. PCR was run in triplicate using the 7500 Fast Real-Time PCR System (Applied Biosystems). SDS v1.4 software (Applied Biosystems) was used for comparative ΔC_T analysis. U6 served as an endogenous control.

DNA extraction, pyrosequencing of KRAS, BRAF *and* PIK3CA *and MSI analysis*

Genomic DNA was extracted from FFPE tissues of colorectal tumours using QIAamp DNA FFPE Tissue Kit (Qiagen). Using extracted genomic DNA, PCR and targeted pyrosequencing were performed for *KRAS* (codons 12 and 13), *BRAF* (*V600E*) and *PIK3CA* (exons 9 and 20) (45). MSI analysis was performed using 10 microsatellite markers, as described previously (46). MSI-high was defined as instability in ≥30% of the markers, and MSI-low/microsatellite stability (MSS) as instability in <30% of the markers (46).

Sodium bisulfite treatment and pyrosequencing to measure MLH1 *promoter methylation*

Bisulfite modification of genomic DNA was performed using a BisulFlash™ DNA Modification Kit (Epigentek, Brooklyn, NY). Bisulfite pyrosequencing for *MLH1* methylation was performed using the PyroMark Kit (Qiagen), as described previously (47).

Colon cancer cell line and miRNA transient transfection

In this study, seven colon cancer cell lines (COLO-320-HSR, DLD-1, HCT-116, HT-29, Lovo, RKO and SW480) were utilized (Supplementary Table 1, available at *Carcinogenesis* Online). Total RNA was extracted from cell pellets using the TRIzol Reagent (Invitrogen by Life Technologies, Carlsbad, CA) according to the manufacturer's instructions. Cells were transfected using the Cell Line Nucleofector Kit V (Lonza, Basel, Switzerland) with a Nucleofector I electroporation device (Lonza) for DLD-1, HCT-116 and RKO and Lipofectamine 2000 (Invitrogen by Life Technologies) for COLO-320-HSR, HT-29, Lovo and SW480, according to the manufacturer's instructions. At 72 h after transfection, the cells were harvested for qRT–PCR or western blotting.

Assays for proliferation and invasion

Proliferation of miRNA transfectants was analysed by measuring the uptake of tritiated thymidine in 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide assay (MTT assay; Sigma–Aldrich, St Louis, MO). In brief, transfected cells were seeded into 96-well plates to a density of 5×10^3 cells per well. After incubation for 0, 24, 48, 72 and 96 h, MTT assays were performed using the Cell Counting Kit-8 (Dojindo, Tokyo, Japan) according to the manufacturer's instructions.

Cell invasion was assessed by a Matrigel invasion assay. After incubation for 24 h, 1×10^6 transfected cells suspended in 500 μl of serum-free medium were added to the top of BD BioCoat Matrigel Invasion Chambers (BD Biosciences, Bedford, MA) prehydrated with phosphate-buffered saline, and 750 μl of medium supplemented with 10% fetal bovine serum was added to the lower wells of the plate. After incubation for 24 h, the invading cells were fixed, stained and analysed under a microscope (Olympus, Tokyo, Japan). Cells were counted in five random fields per membrane. In both assays, the experiments with each cell line were performed three times.

Western blot analysis

Protein expression was analysed using a standard immunoblot procedure with anti-KRAS and anti-BRAF. All primary antibodies were procured from Santa Cruz Biotechnology (Santa Cruz, CA). The anti-β-actin monoclonal antibody was used as a loading control (Oncogene Research Products, La Jolla, CA). The immunoreactive bands were visualized using enhanced chemiluminescence (Thermo Scientific, Rockford, IL).

Statistical analysis

JMP (version 10) and SAS (version 9) software programs were used for statistical analyses (SAS Institute, Cary, NC). All P values were two sided. Univariate analyses were performed to investigate clinicopathological and molecular characteristics according to the miR-31 expression level; a chi-square test or Fisher's exact test was used for categorical data, whereas analysis of variance was used to compare the mean patient age and tumour size. To account for multiple hypothesis testing in associations between miR-31 expression and other 12 covariates, the P value for significance was adjusted by Bonferroni correction to $P = 0.0042$ (=0.05/12).

In survival analysis, the Kaplan–Meier method and log-rank test were used to assess the survival time distribution. Cox proportional hazards regression models were used to compute mortality hazard ratios (HRs) according to the miR-31 expression status. Stratification by the tumour-node-metastasis disease stage (I, IIA, IIB, IIIA, IIIB, IIIC and IV) was performed using the 'strata' option in the SAS 'proc phreg' command. A multivariate model initially included sex (male versus female), age at diagnosis (continuous), tumour size (continuous), year of diagnosis (continuous), tumour location (proximal colon versus distal colon and rectum), tumour differentiation (well to moderate versus poor), MSI status (MSI-high versus MSS/MSI-low), *MLH1* methylation (present versus absent) and mutations of *BRAF*, *KRAS* and *PIK3CA* (present versus absent). A backward elimination was performed with a threshold of $P = 0.10$, to avoid overfitting. Cases with missing information for any of the categorical covariates [tumour differentiation (1.7%), MSI status (1.9%), *MLH1* methylation (4.9%), mutations of *BRAF* (0.1%), *KRAS* (1.5%) and *PIK3CA* (0.1%)] were included in the majority category of the given covariate to avoid overfitting. We confirmed that excluding cases with missing information in any of the covariates did not substantially alter results (data not shown).

A multivariate logistic regression analysis was employed to examine the associations with miR-31 expression status (as an outcome variable), adjusting for potential confounders. The model initially included a similar set of covariates to the initial Cox model. A backward elimination procedure with a threshold of $P = 0.10$ was used to select variables in the final model. Cases with missing information for a given covariate were included in a majority category in the initial model, and if the covariate remained in the final model, those cases were included using a missing indicator variable in the final model. The P value for significance was adjusted by Bonferroni correction to $P = 0.0042$ (=0.05/12).

Results

Detection of high-level miR-31 expression in BRAF*-mutated CRCs on miRNA array analysis*

To examine the miRNA expression signature in *BRAF*-mutated CRCs, 29 cases of CRCs (Supplementary Table 2, available at *Carcinogenesis* Online) were randomly selected from the CRC specimens for miRNA array analysis. Median levels of expression in the *BRAF* mutation group were compared with those in the *BRAF*

wild-type group (including cases with *KRAS* mutation). miRNA array data revealed differential expression in 33 individual miRNAs ($P < 0.05$ by Mann–Whitney *U*-test) between the two groups (Table I). All 33 miRNAs displayed higher expression levels in the *BRAF* mutation group than in the *BRAF* wild-type group. Of the 760 miRNAs, miR-31-5p was up-regulated the most often (335-fold change, $P = 0.009$).

Distribution of miR-31 expression in CRCs and association of miR-31 with clinicopathological and molecular features

We assayed miR-31 expression in 735 FFPE CRC tissue specimens and successfully obtained 721 (98%) valid results. Fourteen patients were unavailable for miR-31 expression analysis because of the lack of extracted RNA from FFPE CRC tissue specimens. We utilized 721 CRC cases, based on the availability of miR-31-5p expression data. miR-31 expression levels were quantified in CRC specimens and paired normal mucosa specimens. miR-31 expression was calculated using the equation $2^{-\Delta C_T}$, where $\Delta C_T = (C_T \text{ miR-31} - C_T \text{U6})$. To calculate the relative expression of miR-31 in each CRC, $2^{-\Delta C_T}$ of cancer tissue was divided by $2^{-\Delta C_T}$ of paired normal tissue. The distributions of miR-31 expression in the 721 CRC specimens were as follows: mean: 41.9; median: 6.3; SD: 176.2; range: 0.04–2108; interquartile range: 2.0–23.4. Cases with miR-31 expression were then divided into quartiles for further analysis: Q1 (<2.0), Q2 (2.0–6.2), Q3 (6.3–23.3) and Q4 (≥23.4). Table II shows the clinicopathological and molecular features of CRCs according to miR-31 expression level. High miR-31 expression was significantly associated with larger tumour size, proximal location, poor differentiation, advanced disease stage, *BRAF* mutation, *KRAS* mutation and MSI-high status ($P \leq 0.0042$ for all).

High miR-31 expression and patient survival

The influence of high miR-31 expression on clinical outcome was assessed in 721 CRC patients (stages I–IV). During follow-up of the 698 patients eligible for survival analysis, mortality occurred in 149, including 115 deaths confirmed to be attributable to CRCs. The median follow-up time for censored patients was 4.7 years. Kaplan–Meier analysis was performed using categorical variables (Q1, Q2, Q3 or Q4). Significantly higher mortality was observed in patients with high miR-31 expression in terms of cancer-specific survival (log-rank test: $P = 0.0013$) and overall survival (log-rank test: $P = 0.0026$) than in those with low miR-31 expression (Figure 1).

In univariate Cox regression analysis, compared with Q1 cases, significantly higher mortality rates were observed in Q2 cases [HR: 1.96; 95% confidence interval (CI): 1.06–3.77; $P = 0.031$], Q3 cases (HR: 2.16; 95% CI: 1.18–4.13; $P = 0.012$) and Q4 cases (HR: 3.10; 95% CI: 1.76–5.78; $P < 0.0001$) (Table III). Similarly, compared with Q1 cases, an independent association with shorter prognosis was observed in Q4 cases in stage-stratified (HR: 2.49; 95% CI: 1.40–4.67; $P = 0.0016$) and multivariate analyses (HR: 2.91; 95% CI: 1.60–5.57; $P = 0.0004$) for cancer-specific survival (Table III). On the other hand, compared with Q1 cases, slightly but insignificantly higher mortality rates were observed in Q2 and Q3 cases in stage-stratified (Q2: $P = 0.11$, Q3: $P = 0.23$) and multivariate stage-stratified analyses (Q2: $P = 0.14$, Q3: $P = 0.18$) (Table III). Similar results were observed in stage-stratified and multivariate stage-stratified analyses for overall survival (data not shown). Therefore, we made a dichotomous miR-31 expression variable, defining Q4 as the 'high-expression group' and combining Q1, Q2 and Q3 into the 'low-expression group'. In multivariate stage-stratified analysis, compared

Table I. Differentially expressed miRNA in *BRAF*-mutated and *BRAF* wild-type CRCs by miRNA array analysis

No.	Name of miRNA (miR base ID)	Relative miRNA expression (miRNA/U6)			*P*
		BRAF mutation group (median; $N = 7$)	*BRAF* wild-type group (median; $N = 22$)	Fold change (mutation group/wild-type group)	
1	hsa-miR-31-5p	29 925.00	89.30	335.0	0.009
2	hsa-miR-215	7.65	0.10	74.5	0.001
3	hsa-miR-151-3p	1312.00	24.40	53.8	0.003
4	hsa-miR-539-5p	370.00	7.57	48.9	0.021
5	hsa-miR-31-3p	77.30	2.14	36.1	0.002
6	hsa-miR-661	3125.00	91.80	34.1	0.011
7	hsa-miR-197-3p	4.78	0.16	29.2	0.002
8	hsa-miR-483-3p	605.00	21.90	27.6	0.032
9	hsa-miR-185-5p	15.40	0.56	27.3	0.024
10	hsa-miR-223-3p	10.50	0.40	25.9	0.005
11	hsa-miR-451a	23.50	1.00	23.5	0.015
12	hsa-miR-410	19.90	0.85	23.4	0.001
13	hsa-miR-15b-5p	9.01	0.40	22.7	0.004
14	hsa-miR-126-5p	8.45	0.37	22.6	0.013
15	hsa-miR-221-3p	99.90	4.82	20.7	0.048
16	hsa-miR-10b-3p	13.80	0.67	20.6	0.015
17	hsa-miR-29c-3p	19.50	1.02	19.1	0.001
18	hsa-miR-625-5p	185.00	10.30	18.0	0.002
19	hsa-miR-34a-5p	75.70	4.37	17.3	0.032
20	hsa-miR-7-1-3p	10.40	0.60	17.3	0.001
21	hsa-miR-10b-5p	23.10	1.49	15.5	0.008
22	hsa-miR-26b-5p	5.46	0.36	15.2	<0.001
23	hsa-let-7a-5p	5.86	0.42	13.9	0.048
24	hsa-miR-145-3p	2.82	0.20	13.8	0.028
25	hsa-miR-374a-5p	26.30	1.96	13.4	0.001
26	hsa-miR-222-3p	11.80	0.89	13.2	0.011
27	hsa-miR-379-5p	1.41	0.11	13.0	0.024
28	hsa-miR-30c-5p	1.59	0.14	11.8	0.002
29	hsa-miR-100-5p	3.26	0.28	11.6	0.004
30	hsa-miR-625-3p	13.20	1.16	11.4	0.001
31	hsa-miR-142-5p	5.23	0.48	11.0	0.002
32	hsa-miR-99a-5p	2.02	0.19	10.5	0.003
33	hsa-miR-425-5p	10.80	1.07	10.1	0.005

The fold change is expressed as the median of the *BRAF* mutation group divided by that of the wild-type group for each miRNA. *P* values were determined by the Mann–Whitney *U*-test. miRNAs with $P < 0.05$ are listed.

Table II. Clinicopathological and molecular features of 721 CRCs according to quartiles of miR-31 expression

Clinicopathological or molecular feature	Total *N*	miR-31 expression				*P*
		Q1 (<2.0)	Q2 (2.0–6.2)	Q3 (6.3–23.3)	Q4 (≥23.4)	
All cases	721	180	180	181	180	
Gender						
Male	422 (59%)	101 (56%)	107 (59%)	103 (57%)	111 (62%)	0.70
Female	299 (41%)	79 (44%)	73 (41%)	78 (43%)	69 (38%)	
Age (mean ± SD)	66.9±11.4	65.7±11.3	67.3±11.0	67.8±12.0	66.7±11.2	0.34
Tumour size (mm) (mean ± SD)	46.5±23.6	36.6±17.3	44.8±20.9	50.0±23.1	54.4±27.9	<0.0001
Year of diagnosis						
Prior to 2002	334 (46%)	93 (52%)	80 (44%)	76 (42%)	85 (47%)	0.29
2003–2012	387 (54%)	87 (48%)	100 (56%)	105 (58%)	95 (53%)	
Tumour location						
Rectum and distal colon (splenic flexure to sigmoid)	465 (64%)	135 (75%)	125 (69%)	119 (66%)	86 (48%)	<0.0001
Proximal colon (caecum to transverse)	256 (36%)	45 (25%)	55 (31%)	62 (34%)	94 (52%)	
Tumour differentiation						
Well to moderate	655 (92%)	176 (99%)	170 (96%)	162 (91%)	147 (84%)	<0.0001
Poor	54 (8.0%)	2 (1.1%)	7 (4.0%)	16 (9.0%)	29 (16%)	
Disease stage						
I	138 (19%)	59 (32%)	31 (17%)	23 (13%)	25 (14%)	<0.0001
IIA	157 (22%)	46 (26%)	42 (23%)	37 (20%)	32 (18%)	
IIB	57 (7.9%)	10 (5.6%)	10 (5.6%)	20 (11%)	17 (9.4%)	
IIIA	41 (5.7%)	14 (7.8%)	11 (6.1%)	12 (6.6%)	4 (2.2%)	
IIIB	163 (23%)	23 (13%)	45 (25%)	46 (25%)	49 (27%)	
IIIC	82 (11%)	14 (7.8%)	20 (11%)	21 (12%)	27 (15%)	
IV	83 (12%)	14 (7.8%)	21 (12%)	22 (12%)	26 (14%)	
BRAF mutation						
Wild-type	685 (95%)	176 (98%)	179 (99%)	173 (96%)	157 (87%)	<0.0001
Mutant	35 (4.9%)	3 (1.7%)	1 (0.6%)	8 (4.4%)	23 (13%)	
KRAS mutation						
Wild-type	479 (67%)	130 (74%)	123 (69%)	126 (70%)	100 (57%)	0.0042
Mutant	231 (33%)	45 (26%)	56 (31%)	54 (30%)	76 (43%)	
PIK3CA mutation						
Wild-type	642 (89%)	168 (94%)	160 (89%)	163 (90%)	151 (84%)	0.023
Mutant	78 (11%)	11 (6.2%)	20 (11%)	18 (9.9%)	29 (16%)	
MSI status						
MSS/MSI-low	658 (93%)	172 (99%)	171 (97%)	164 (92%)	151 (85%)	<0.0001
MSI-high	49 (6.9%)	2 (1.2%)	6 (3.4%)	14 (7.9%)	27 (15%)	
MLH1 methylation						
Unmethylated	387 (56%)	106 (62%)	102 (59%)	94 (54%)	85 (51%)	0.15
Methylated	299 (44%)	65 (38%)	71 (41%)	80 (46%)	83 (49%)	

Percentage (%) indicates the proportion of cases with a specific clinicopathological or molecular feature within a given quartile category (Q1, Q2, Q3 or Q4) of miR-31 expression by qRT–PCR. *P* values were calculated by analysis of variance for age and tumour size and by a chi-square test or Fisher's exact test for all other variables. To account for multiple hypothesis testing in associations between miR-31 expression and other 12 covariates, the *P* value for significance was adjusted by Bonferroni correction to $P = 0.0042$ (=0.05/12).

with the 'low-expression group', a significantly higher mortality rate was observed in the 'high-expression group' (HR: 2.06; 95% CI: 1.36–3.09; $P = 0.0008$) in cancer-specific analysis (Table III).

In stage-stratified (stages I–IV) analysis, the mortality rate in terms of cancer-specific survival was significantly higher in CRC groups (stages II–IV) with high miR-31 expression levels (log-rank test: $P = 0.035$, $P = 0.020$ and $P = 0.024$, respectively) than in those with low miR-31 expression levels (Supplementary Figure 1, available at *Carcinogenesis* Online). Our data also showed that high miR-31 expression was related to cancer-specific mortality, regardless of *BRAF* status (Supplementary Figure 2, available at *Carcinogenesis* Online).

Multivariate logistic regression analysis in cases of high miR-31 expression

Considering potential confounding and potential cause-effect sequence, we performed a multivariate logistic regression analysis to assess the relationships with miR-31 expression. The results showed that high miR-31 (Q4) expression was significantly associated with *BRAF* [odds ratio (OR): 7.05; 95% CI: 3.08–16.8; $P < 0.0001$], *KRAS* mutation (OR: 2.61; 95% CI: 1.75–3.90; $P < 0.0001$) and tumour location in the proximal colon (OR: 2.21; 95% CI: 1.49–3.28; $P < 0.0001$) (Table IV).

Association of miR-31 expression and clinicopathological and molecular features in serrated lesions

We assessed 650 FFPE tissue specimens of serrated lesions and non-serrated adenomas in the miR-31 expression assay and successfully obtained 632 (97%) valid results. Then miR-31 expression levels were also quantified in 381 colorectal serrated lesions and 251 non-serrated adenomas. It is very difficult to obtain miRNA specimens of paired normal tissue for comparison with colorectal serrated lesions and non-serrated adenomas by endoscopic resection

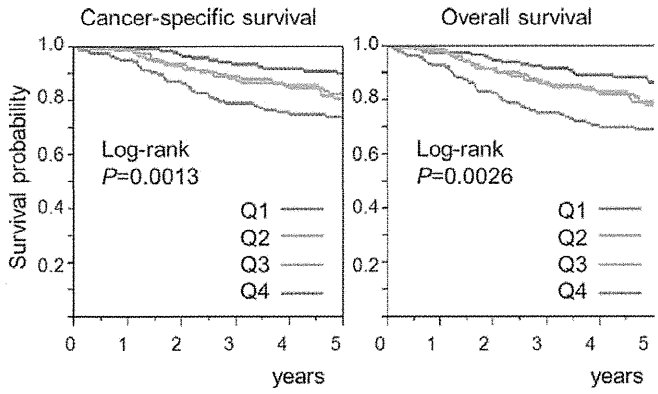

Fig. 1. Kaplan–Meier survival curves for CRCs (stages I–IV; $N = 698$) according to the miR-31 expression level. Cancer-specific survival: left panel and overall survival: right panel. Significantly higher mortality rates were observed in patients with high miR-31 expression than in those with low expression for both cancer-specific (log-rank test: $P = 0.0013$) and overall survival (log-rank test: $P = 0.0026$).

because of the small size of the resected sample. Therefore, we incorporated pooled normal mucosa specimens in each plate to standardize all assay runs. Distributions of miR-31 expression were as follows (mean ± SD; median): HP (10.3 ± 19.3; 2.6), SSA/P (20.5 ± 25.1; 12.9), TSA (24.8 ± 28.0; 14.7) and non-serrated adenoma (14.1 ± 24.9; 3.4).

Supplementary Table 3, available at *Carcinogenesis* Online, shows the clinicopathological and molecular features, including miR-31 expression, in serrated lesions and non-serrated adenomas. High miR-31 expression (Q4 in CRCs; expression level ≥23.4) was frequently detected in cases of SSA/P [32% (41/128)] and TSA [37% (42/113)] compared with those of HP [13% (18/140)] and non-serrated adenoma [19% (45/251)] ($P < 0.0001$). Multivariate regression analysis was adjusted for potential confounders including *BRAF* and *KRAS* mutations, tumour location and tumour size. The results showed a persistent significant association between high miR-31 expression and histological type [SSA/P: $P = 0.0092$, TSA: $P < 0.0001$ (HP as a referent)].

Functional analysis of miR-31 expression in colon cancer cell lines

miR-31 mimics and the inhibitor were transfected into colon cancer cell lines. The results confirmed the up-regulation or down-regulation of miR-31 expression (Supplementary Figure 3, available at *Carcinogenesis* Online). The Matrigel invasion assay revealed enhanced invasive potential of the miR-31 mimic (Figure 2A) after transfection (72 h later) into cancer cell lines. Similar results were observed in the proliferation assay (data not shown). In the proliferation assay, significantly decreased cell proliferation was also observed as a result of transfection (96 h later) of the miR-31 inhibitor (Figure 2B). Similar results were observed in the invasion assay (data not shown).

To determine the effect of miR-31 on *BRAF* and *KRAS* target proteins, expression of those proteins was compared before and after transfection (72 h later) of the miR-31 inhibitor into the cell lines. The results of western blot analysis demonstrated that after transfection, *BRAF* target proteins decreased in colon cancer cell lines, regardless of the mutational status (Figure 2C). In contrast, none of the colon cancer cell lines showed a decrease in *KRAS* target proteins.

Discussion

In this study, specific miRNA expression associated with *BRAF* (*V600E*) mutation was identified. The results of miRNA array analysis revealed that miR-31 was the most up-regulated gene in *BRAF*-mutated CRCs compared with *BRAF* wild-type CRCs. In a database of 721 patients with CRC, high miR-31 expression was associated with *BRAF* and *KRAS* mutations and proximal location in multivariate logistic regression analysis. After the transfection of the miR-31 inhibitor, western blot analysis revealed a decrease in BRAF target protein in colon cancer cell line. Thus, our data support the hypothesis

Table III. Association of miR-31 expression with patient mortality in CRCs

miR-31 expression (quartile)	Total *N*	Cancer-specific survival		
		Univariate	Stage-stratified	Multivariate stage-stratified
		HR (95% CI)	HR (95% CI)	HR (95% CI)
Q1 (<2.0)	171	1 (referent)	1 (referent)	1 (referent)
Q2 (2.0–6.2)	174	1.96 (1.06–3.77)	1.65 (0.89–3.18)	1.59 (0.86–3.07)
Q3 (6.3–23.9)	177	2.16 (1.18–4.13)	1.46 (0.79–2.81)	1.53 (0.83–2.96)
Q4 (≥24.0)	176	3.10 (1.76–5.78)	2.49 (1.40–4.67)	2.91 (1.60–5.57)
P for trend		0.0009	0.013	0.0032
Low-expression group (Q1–3)	522	1 (referent)	1 (referent)	1 (referent)
High-expression group (Q4)	176	1.84 (1.25–2.67)	1.78 (1.20–2.60)	2.06 (1.36–3.09)
P		0.0024	0.0043	0.0008

The multivariate, stage-stratified Cox model included the miR-31 expression variable stratified by sex, age at diagnosis, tumour size, year of diagnosis, tumour location, tumour differentiation, MSI status, *MLH1* methylation and mutations of *BRAF*, *KRAS* and *PIK3CA*.

Table IV. Multivariate logistic regression analysis of miR-31 expression in CRCs

Variables in the final model for miR-31 expression (as an outcome variable) [high-expression group (Q4) versus low-expression group (Q1–3)]	Adjusted OR (95% CI)	*P*
BRAF mutant (versus wild-type)	7.05 (3.08–16.8)	<0.0001
KRAS mutant (versus wild-type)	2.61 (1.75–3.90)	<0.0001
Proximal colon (versus distal colon and rectum)	2.21 (1.49–3.28)	<0.0001
Poor differentiation (versus well to moderate)	2.75 (1.38–5.44)	0.0044
Tumour size (for 30 mm increase as a unit)	1.46 (1.11–1.92)	0.0062

A multivariate logistic regression analysis assessing the relationships with miR-31 expression status initially included sex, age, tumour size, year of diagnosis, tumour location, tumour differentiation, disease stage, MSI, *MLH1* methylation, and mutations of *BRAF*, *KRAS* and *PIK3CA*, considering potential confounding and causal relationships. For multiple hypothesis testing, the *P* value for significance was adjusted by Bonferroni correction to 0.0042 (=0.05/12).

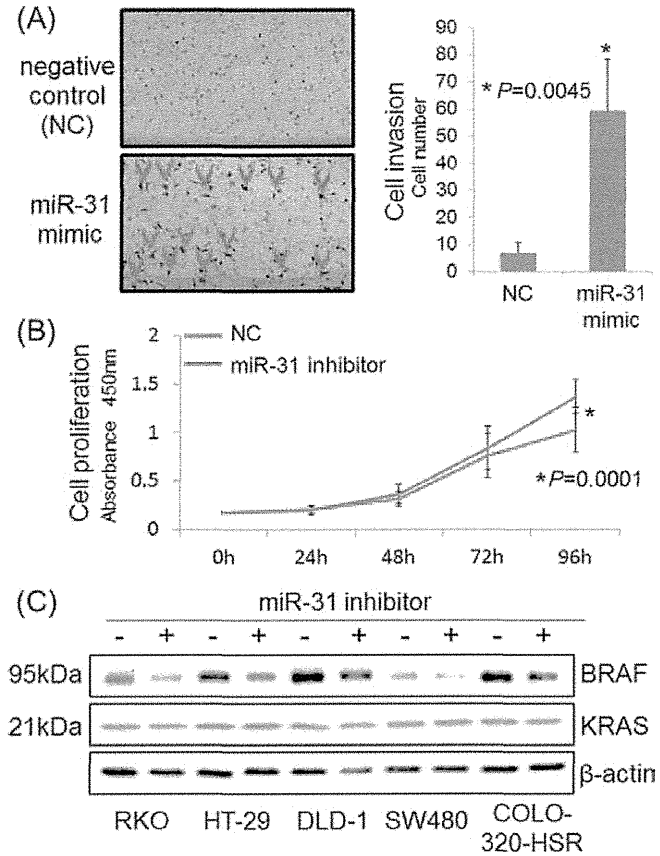

Fig. 2. Functional analysis of miR-31. (**A**) Results of the Matrigel invasion assay. Invading cells are indicated by arrow heads in the left panel. The right panel represents the means of five random microscopic fields per membrane; error bars represent the standard deviations. This assay revealed that the miR-31 mimic enhanced invasion by 8.5-fold ($P = 0.0045$) in SW480 cells (*KRAS* mutated) after transfection (72 h later). The P value was analysed using a paired T-test. (**B**) In the proliferation assay, the miR-31 inhibitor significantly decreased cell proliferation in HT-29 cells (*BRAF* mutated) ($P = 0.0001$). The graph depicts the means of 16 replications; error bars represent standard deviations. (**C**) In western blot analysis, after transfection (72 h later) of the miR-31 inhibitor, *BRAF* target proteins were decreased in RKO cells (*BRAF* mutated), HT-29 cells (*BRAF* mutated), DLD-1 cells (*KRAS* mutated), SW480 cells (*KRAS* mutated) and COLO-320-HSR (wild-type), respectively.

that miR-31 may regulate *BRAF* activation in CRCs. We also identified that high miR-31 expression was an unfavourable prognostic factor in patients with CRC, independent of clinicopathological and molecular features. In contrast, high miR-31 expression was frequently detected in cases with SSA/P and TSA compared with those with HP, suggesting an oncogenic role of this miRNA in the serrated pathway. The transfection of the miR-31 inhibitor exhibited an antitumour effect in functional analysis. Therefore, miR-31 may be a promising diagnostic biomarker and the therapeutic target in patients with CRC.

miR-31 is located at 9p21.3 and is reportedly deregulated in various human cancers (8,9,11,12,15). Previous studies have shown that miR-31 has oncogenic potential in oesophageal squamous cell carcinoma (8) and acts as a tumour suppressor in oesophageal adenocarcinoma (15), gastric cancer (12), ovarian cancer (11) and breast cancer (9). With regard to CRC, an association has been reported between miR-31, oncogenic potential (21–24,26,48), deeper invasion (24,48) and advanced disease stage (21,24,48); however, none of these studies have examined the association between miR-31 expression and mortality in CRC patients. In the present study, a large database was utilized. High miR-31 expression was independently associated with shorter prognosis in the multivariate stage-stratified Cox model. The importance of large-scale studies cannot be emphasized enough. Small studies with null results are much less likely to remain unpublished compared with small studies with significant results; this leads to publication bias. In contrast to previous studies (21–24,26,48), the present study examined the miR-31 expression status in a much larger sample of CRCs. Therefore, our data support the hypothesis that high miR-31 expression may be a prognostic biomarker of CRC.w Nevertheless, the data on CRCs presented here have some limitations, including the cross-sectional, observational nature of the study. Future independent studies should confirm the correlation between miR-31 and unfavourable prognosis in patients with CRC.

The tumour molecular characterization for personalized medicine is becoming important in CRCs (1–5,16,17,45). Accumulating evidence suggests that similar to *PIK3CA* and *PTEN* mutations, *BRAF* mutations confer therapeutic resistance to cetuximab and panitumumab (2,3) in patients with CRCs because these genes are located downstream of EGFR. In addition, a relationship between *BRAF* mutation and unfavourable survival has been previously reported in patients with CRCs (2,5). These results suggest that *BRAF* mutation can be a new biomarker for molecular diagnosis and identification of prognostic factors; however, no previous study has identified specific miRNAs associated with *BRAF* mutation in a large sample of colorectal tumours. In the current study, associations were identified between miR-31 expression and *BRAF* mutation and CRC prognosis. Previous studies have detected high miR-31 expression in CRC patients with MSI-high status (20,28,34) or poor differentiation (21,23). In the present study, high miR-31 expression was significantly associated with MSI-high status in univariate analysis; however, no significant association was observed in multivariate analysis. Furthermore, in serrated lesions, despite the fact that MSI-high was quite low [1.5% (6/381)], high miR-31 expression was frequently detected in cases with SSA/P and TSA. Thus, high miR-31 expression in CRC patients with MSI-high in previous studies (20,28,34) may have been due to *BRAF* mutation, which has been strongly associated with MSI-high status (1,5).

Recent studies have reported that several miRNAs target the genes in the downstream effectors of EGFR, such as miR-143 and miR-145 (25,29) for *KRAS*, miR-520a and miR-525 (30) for *PIK3CA* and miR-21 (31,32) and miR-155 (32) for *PTEN*. Moreover, *BRAF* is thought to be targeted by miR-143 and miR-145, which play a role as tumour suppressors (29). In the present study, high miR-31 expression was strongly associated with *BRAF* mutation in a CRC large sample. In addition, after transfection of the miR-31 inhibitor, western blot analysis revealed a decrease in *BRAF* target protein. These results support the hypothesis that miR-31 may regulate the activation of *BRAF* gene in CRC. The exact mechanism of this regulation by miR-31 remains unknown; however, a recent study has reported that miR-31 may target a RAS p21 GTPase-activating protein 1 (RASA1), which is a negative regulator of the RAS–RAF–MEK–ERK signalling pathway (35). Therefore, miR-31 may regulate *BRAF* activity via suppression of RASA1 in CRC, resulting in up-regulation of the signalling pathway. These findings also imply that miR-31 may serve as a molecular target of the RAF or MEK inhibitor.

Our data also showed a decrease in *BRAF* target proteins regardless of the mutational status after transfection of the miR-31 inhibitor. This decrease in *BRAF* target proteins was observed in all cell lines; however, none of the colon cancer cell lines exhibited a decrease in *KRAS* target protein. One possible explanation for these phenomena is that miR-31 may target the negative regulator, which plays a role in the pathways downstream of RAS. Further functional analysis is required to clarify the regulatory role of miR-31 in the RAS–RAF–MEK–ERK signalling pathway and its potential as a molecular target of those inhibitors.

Previously, SSA/P was often classified as HP, which was considered to have no malignant potential. However, recent studies have shown that SSA/P is mainly observed in the proximal colon (49) and is associated with frequent *BRAF* mutation and *MLH1* methylation (40–42). These results suggest that SSA/P possesses malignant potential and might be a precursor lesion of MSI-high CRC with *BRAF* mutation in the proximal colon. With regard to miRNA expression in serrated lesions, a previous study involving serrated lesions ($N = 37$) reported that SSA/P was characterized by high levels of miR-181b

and miR-21 expression compared with HP (50). However, the authors concluded that discrimination between the two lesions was impossible on the basis of miR-181b and miR-21 expression. Thus, the effects of miRNA expression in serrated lesions remain largely unknown. In the present study, high miR-31 expression was frequently detected in cases with SSA/P and TSA compared with those with HP in large samples of serrated lesions ($N = 381$). After adjusting for *BRAF* and *KRAS* mutation status, tumour location and tumour size, a persistently significant association between high miR-31 expression and the pathological features of SSA/P and TSA was observed. Thus, our data suggest that high miR-31 expression may occur in the early stage of colorectal tumorigenesis and play an oncogenic role in serrated lesions. Moreover, our findings challenge the common conception of discrete molecular features of SSA/Ps versus HPs. They may, therefore, have a substantial impact on clinical and translational research.

In conclusion, in this study, high miR-31 expression was associated with *BRAF* mutation involving a CRC large sample. This result may indicate that miR-31 is one of the important miRNAs in CRC with *BRAF* mutation. In addition, high miR-31 expression was associated with patient mortality. Finally, an antitumour effect was observed as a result of transfection of the miR-31 inhibitor. Thus, miR-31 may be a promising diagnostic biomarker and therapeutic target in patients with CRC. Moreover, our data suggest that miR-31 may play an important role in the progression of serrated lesions.

Supplementary material

Supplementary Tables 1–3 and Figures 1–3 can be found at http://carcin.oxfordjournals.org/

Funding

Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (23790800); A-STEP (Adaptable & Seamless Technology Transfer Program through Target-driven R&D); Daiwa Securities Health Foundation; Kobayashi Foundation for Cancer Research; Sagawa Foundation for Promotion of Cancer Research; Suzuken memorial foundation and Takeda Science Foundation.

Acknowledgements

We deeply thank the pathology departments of Sapporo Medical University Hospital, Keiyukai Sapporo Hospital and JR Sapporo Hospital for providing us with tissue specimens. The authors would like to thank Enago (www.enago.jp) for the English language review. Study concept and design: K.N.; acquisition of data: K.N., H.I., M.I., T.N., W.S., M.F. and T.H.; analysis and interpretation of data: K.N., H.I., R.M.M.N., H.S., H.Y. and Y.S.; drafting of the manuscript: K.N., H.I., M.I. and H.Y.; critical revision of the manuscript for important intellectual content: K.N., H.I., M.I., R.M., Y.S., E.Y., S.K. and Y.S.; statistical analysis: K.N. and M.N.; material support: S.Y., M.M., H.T., H.O., T.K., M.H., M.F., T.H., K.O. and K.H.; study supervision: H.S., H.Y. and Y.S.; final approval of manuscript: all authors.

Conflict of Interest Statement: None declared.

References

1. Network,T.C.G.A. (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, **487**, 330–337.
2. De Roock,W. *et al.* (2011) KRAS, BRAF, PIK3CA, and PTEN mutations: implications for targeted therapies in metastatic colorectal cancer. *Lancet Oncol.*, **12**, 594–603.
3. Di Nicolantonio,F. *et al.* (2008) Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer. *J. Clin. Oncol.*, **26**, 5705–5712.
4. Donehower,L.A. *et al.* (2013) MLH1-silenced and non-silenced subgroups of hypermutated colorectal carcinomas have distinct mutational landscapes. *J. Pathol.*, **229**, 99–110.
5. Lochhead,P. *et al.* (2013) Microsatellite instability and BRAF mutation testing in colorectal cancer prognostication. *J. Natl. Cancer Inst*, **105**, 1151–1156.
6. Nishikawa,E. *et al.* (2011) miR-375 is activated by ASH1 and inhibits YAP1 in a lineage-dependent manner in lung cancer. *Cancer Res.*, **71**, 6165–6173.
7. Lu,J. *et al.* (2005) MicroRNA expression profiles classify human cancers. *Nature*, **435**, 834–838.
8. Zhang,T. *et al.* (2011) The oncogenetic role of microRNA-31 as a potential biomarker in oesophageal squamous cell carcinoma. *Clin. Sci. (Lond).*, **121**, 437–447.
9. Valastyan,S. *et al.* (2009) A pleiotropically acting microRNA, miR-31, inhibits breast cancer metastasis. *Cell*, **137**, 1032–1046.
10. Meng,F. *et al.* (2007) MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer. *Gastroenterology*, **133**, 647–658.
11. Creighton,C.J. *et al.* (2010) Molecular profiling uncovers a p53-associated role for microRNA-31 in inhibiting the proliferation of serous ovarian carcinomas and other cancers. *Cancer Res.*, **70**, 1906–1915.
12. Zhang,Y. *et al.* (2010) Down-regulation of miR-31 expression in gastric cancer tissues and its clinical significance. *Med. Oncol.*, **27**, 685–689.
13. Schaefer,A. *et al.* (2010) Diagnostic and prognostic implications of microRNA profiling in prostate carcinoma. *Int. J. Cancer*, **126**, 1166–1176.
14. Ueda,T. *et al.* (2010) Relation between microRNA expression and progression and prognosis of gastric cancer: a microRNA expression analysis. *Lancet Oncol.*, **11**, 136–146.
15. Leidner,R.S. *et al.* (2012) The microRNAs, MiR-31 and MiR-375, as candidate markers in Barrett's esophageal carcinogenesis. *Genes. Chromosomes Cancer*, **51**, 473–479.
16. Bartley,A.N. *et al.* (2011) Complex patterns of altered MicroRNA expression during the adenoma-adenocarcinoma sequence for microsatellite-stable colorectal cancer. *Clin. Cancer Res.*, **17**, 7283–7293.
17. Balaguer,F. *et al.* (2011) Colorectal cancers with microsatellite instability display unique miRNA profiles. *Clin. Cancer Res.*, **17**, 6239–6249.
18. Schetter,A.J. *et al.* (2008) MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *JAMA*, **299**, 425–436.
19. Shibuya,H. *et al.* (2010) Clinicopathological and prognostic value of microRNA-21 and microRNA-155 in colorectal cancer. *Oncology*, **79**, 313–320.
20. Sarver,A.L. *et al.* (2009) Human colon cancer profiles show differential microRNA expression depending on mismatch repair status and are characteristic of undifferentiated proliferative states. *BMC Cancer*, **9**, 401.
21. Schee,K. *et al.* (2012) Clinical relevance of microRNA miR-21, miR-31, miR-92a, miR-101, miR-106a and miR-145 in colorectal cancer. *BMC Cancer*, **12**, 505.
22. Cekaite,L. *et al.* (2012) MiR-9, -31, and -182 deregulation promote proliferation and tumor cell survival in colon cancer. *Neoplasia*, **14**, 868–879.
23. Chang,K.H. *et al.* (2011) MicroRNA signature analysis in colorectal cancer: identification of expression profiles in stage II tumors associated with aggressive disease. *Int. J. Colorectal Dis.*, **26**, 1415–1422.
24. Wang,C.J. *et al.* (2009) Clinicopathological significance of microRNA-31, -143 and -145 expression in colorectal cancer. *Dis. Markers*, **26**, 27–34.
25. Chen,X. *et al.* (2009) Role of miR-143 targeting KRAS in colorectal tumorigenesis. *Oncogene*, **28**, 1385–1392.
26. Slaby,O. *et al.* (2007) Altered expression of miR-21, miR-31, miR-143 and miR-145 is related to clinicopathologic features of colorectal cancer. *Oncology*, **72**, 397–402.
27. Bandrés,E. *et al.* (2006) Identification by Real-time PCR of 13 mature microRNAs differentially expressed in colorectal cancer and non-tumoral tissues. *Mol. Cancer*, **5**, 29.
28. Earle,J.S. *et al.* (2010) Association of microRNA expression with microsatellite instability status in colorectal adenocarcinoma. *J. Mol. Diagn.*, **12**, 433–440.
29. Pagliuca,A. *et al.* (2013) Analysis of the combined action of miR-143 and miR-145 on oncogenic pathways in colorectal cancer cells reveals a coordinate program of gene repression. *Oncogene*, **32**, 4806–4813.
30. Arcaroli,J.J. *et al.* (2012) Common PIK3CA mutants and a novel 3' UTR mutation are associated with increased sensitivity to saracatinib. *Clin. Cancer Res.*, **18**, 2704–2714.
31. Xiong,B. *et al.* (2013) MiR-21 regulates biological behavior through the PTEN/PI-3 K/Akt signaling pathway in human colorectal cancer cells. *Int. J. Oncol.*, **42**, 219–228.
32. Bakirtzi,K. *et al.* (2011) Neurotensin signaling activates microRNAs-21 and -155 and Akt, promotes tumor growth in mice, and is increased in human colon tumors. *Gastroenterology*, **141**, 1749–61.e1.

33. Mosakhani,N. *et al.* (2012) MicroRNA profiling differentiates colorectal cancer according to KRAS status. *Genes. Chromosomes Cancer*, **51**, 1–9.
34. Oberg,A.L. *et al.* (2011) miRNA expression in colon polyps provides evidence for a multihit model of colon cancer. *PLoS One*, **6**, e20465.
35. Sun,D. *et al.* (2013) MicroRNA-31 activates the RAS pathway and functions as an oncogenic MicroRNA in human colorectal cancer by repressing RAS p21 GTPase activating protein 1 (RASA1). *J. Biol. Chem.*, **288**, 9508–9518.
36. Hur,K. *et al.* (2013) MicroRNA-200c modulates epithelial-to-mesenchymal transition (EMT) in human colorectal cancer metastasis. *Gut*, **62**, 1315–1326.
37. Takahashi,M. *et al.* (2012) Boswellic acid exerts antitumor effects in colorectal cancer cells by modulating expression of the let-7 and miR-200 microRNA family. *Carcinogenesis*, **33**, 2441–2449.
38. Wu,J. *et al.* (2012) MicroRNA-34a inhibits migration and invasion of colon cancer cells via targeting to Fra-1. *Carcinogenesis*, **33**, 519–528.
39. Yu,Y. *et al.* (2012) MicroRNA-21 induces stemness by downregulating transforming growth factor beta receptor 2 (TGFβR2) in colon cancer cells. *Carcinogenesis*, **33**, 68–76.
40. Bettington,M. *et al.* (2013) The serrated pathway to colorectal carcinoma: current concepts and challenges. *Histopathology*, **62**, 367–386.
41. Leggett,B. *et al.* (2010) Role of the serrated pathway in colorectal cancer pathogenesis. *Gastroenterology*, **138**, 2088–2100.
42. Rosty,C. *et al.* (2012) Phenotype and polyp landscape in serrated polyposis syndrome: a series of 100 patients from genetics clinics. *Am. J. Surg. Pathol.*, **36**, 876–882.
43. McShane,L.M. *et al.*; Statistics Subcommittee of the NCI-EORTC Working Group on Cancer Diagnostics. (2005) Reporting recommendations for tumor marker prognostic studies (REMARK). *J. Natl. Cancer Inst.*, **97**, 1180–1184.
44. Bosman,F.T. *et al.* (2010) *WHO Classification of Tumours of the Digestive System*. International Agency for Research on Cancer, Lyon, France.
45. Liao,X. *et al.* (2012) Aspirin use, tumor PIK3CA mutation, and colorectal-cancer survival. *N. Engl. J. Med.*, **367**, 1596–1606.
46. Nosho,K. *et al.* (2009) A prospective cohort study shows unique epigenetic, genetic, and prognostic features of synchronous colorectal cancers. *Gastroenterology*, **137**, 1609–20.e1.
47. Iwagami,S. *et al.* (2013) LINE-1 hypomethylation is associated with a poor prognosis among patients with curatively resected esophageal squamous cell carcinoma. *Ann. Surg.*, **257**, 449–455.
48. Cottonham,C.L. *et al.* (2010) miR-21 and miR-31 converge on TIAM1 to regulate migration and invasion of colon carcinoma cells. *J. Biol. Chem.*, **285**, 35293–35302.
49. Torlakovic,E.E. *et al.* (2008) Sessile serrated adenoma (SSA) vs. traditional serrated adenoma (TSA). *Am. J. Surg. Pathol.*, **32**, 21–29.
50. Schmitz,K.J. *et al* (2009) Differential expression of microRNA 181b and microRNA 21 in hyperplastic polyps and sessile serrated adenomas of the colon. *Virchows Arch.*, **455**, 49–54.

Received September, 4, 2013; revised October 16, 2013; accepted November 9, 2013

Hepatic biliary epithelial cells acquire epithelial integrity but lose plasticity to differentiate into hepatocytes *in vitro* during development

Naoki Tanimizu[1,*], Yukio Nakamura[2], Norihisa Ichinohe[1], Toru Mizuguchi[2], Koichi Hirata[2] and Toshihiro Mitaka[1]

[1]Department of Tissue Development and Regeneration, Research Institute for Frontier Medicine, Sapporo Medical University School of Medicine, S-1, W-17, Chuo-ku, Sapporo, Japan
[2]First Department of Surgery, Sapporo Medical University School of Medicine, Sapporo Medical University School of Medicine, S-1, W-17, Chuo-ku, Sapporo, Japan
*Author for correspondence (tanimizu@sapmed.ac.jp)

Accepted 4 September 2013
Journal of Cell Science 126, 5239 5246
© 2013. Published by The Company of Biologists Ltd
doi: 10.1242/jcs.133082

Summary

In developing organs, epithelial tissue structures are mostly developed by the perinatal period. However, it is unknown whether epithelial cells are already functionally mature and whether they are fixed in their lineage. Here we show that epithelial cells alter their plasticity during postnatal development by examining the differentiation potential of epithelial cell adhesion molecule $(EpCAM)^+$ cholangiocytes (biliary epithelial cells) isolated from neonatal and adult mouse livers. We found that neonatal cholangiocytes isolated from 1 week old liver converted into functional hepatocytes in the presence of oncostatin M and Matrigel®. In contrast, neither morphological changes nor expression of hepatocyte markers were induced in adult cholangiocytes. The transcription factors hepatocyte nuclear factor 4α and CCAAT/enhancer binding protein α (C/EBPα), which are necessary for hepatocytic differentiation, were induced in neonatal cholangiocytes but not in adult cells, whereas grainyhead like 2 (Grhl2) and hairy enhance of slit 1 (Hes1), which are implicated in cholangiocyte differentiation, were continuously expressed in adult cells. Overexpression of C/EBPα and Grhl2 promoted and inhibited hepatocytic differentiation, respectively. Furthermore, adult cholangiocytes formed a monolayer with higher barrier function than neonatal ones did, suggesting that cholangiocytes are still in the process of epithelial maturation even after forming tubular structures during the neonatal period. Taken together, these results suggest that cholangiocytes lose plasticity to convert into hepatocytes during epithelial maturation. They lose competency to upregulate hepatocytic transcription factors and downregulate cholagiocytic ones under conditions inducing hepatocytic differentiation. Our results suggest that a molecular machinery augmenting epithelial integrity limits lineage plasticity of epithelial cells.

Key words: Epithelial progenitors, Plasticity, Cholangiocytes, Bile duct, Mature hepatocytes

Introduction

During development, tissue stem/progenitor cells differentiate to multiple types of epithelial cells, which establish various tissue structures, including alveoli in the lung, renal tubules in the kidney, and hepatic cords and bile ducts in the liver. Given that organs need to perform their physiological functions after the birth, epithelial tissue structures may be mostly developed at the birth or soon after. However, it is unknown whether epithelial cells are fixed in their lineage and fully functional in neonatal organs.

The liver contains two types of epithelial cells, named hepatocytes and cholangiocytes, which originate from hepatoblasts (fetal liver stem/progenitor cells) during development (Oertel et al., 2003; Tanimizu et al., 2003). Cholangiocytes are biliary epithelial cells forming bile duct tubules. Bile ducts connect the liver to the intestine to drain the bile secreted by hepatocytes. It can be assumed that cholangiocytes acquire epithelial characteristics including secretory and barrier functions when they establish the tubular structure, since it is physiologically important to modulate the composition of bile and avoid any leakage of bile during drainage. However, it is unknown whether cholangiocytes in the neonatal liver have similar epithelial characteristics as those in the adult liver.

In the adult liver, there are at least three possible sources of hepatocytes and cholangiocytes: self duplication of mature cells, the stem cell system, and lineage conversion. The self duplication of hepatocytes and cholangiocytes to replace aged or damaged cells is the simplest way, which may be the case in normal and in acutely injured livers (Michalopoulos, 2007; Malato et al., 2011). In contrast, after severe chronic liver injury, the duplication ability of the epithelial cells may be exhausted and stem or progenitor cells may be activated to supply hepatocytes and cholangiocytes (Espanol Suner et al., 2012). In addition to the self duplication and stem/progenitor cell systems, lineage conversion should be taken into consideration (Michalopoulos, 2011). It has been shown that mature hepatocytes (MHs) have the potential to transdifferentiate into cholangiocyte like cells (Nishikawa et al., 2005; Zong et al., 2009). In contrast to hepatocytes, it remains unclear whether cholangiocytes have the ability to convert into hepatocytes.

Journal of Cell Science