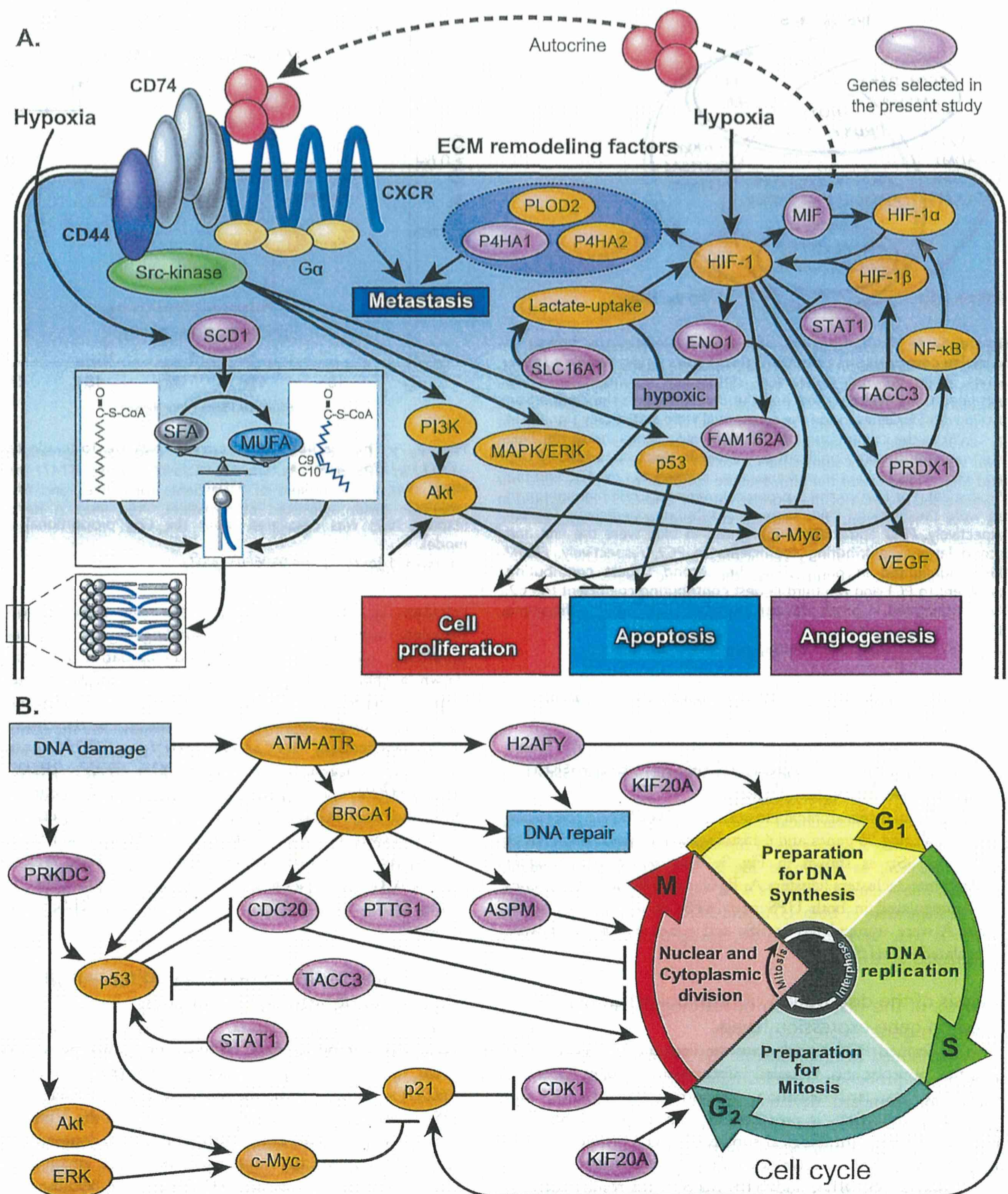


Figure 8. A hypothetical regulation model of metabolic and signaling control in highly malignant STS. (A) Signaling pathways, excluding cell cycle and DNA repair. (B) Cell cycle and DNA repair pathways. The pink oval indicates the genes selected in the present study. MUFA, monounsaturated fatty acid; SFA, saturated fatty acid; SCD1, stearoyl-CoA desaturase 1; MIF, macrophage migration inhibitory factor; CXCR, CXC chemokine receptor; PI3K, phosphoinositide 3-kinase; MAPK, extracellular signal-regulated kinase; ERK, mitogen-activated protein kinase; PTTG1, pituitary tumor-transforming 1; ASPM, abnormal spindle-like microcephaly-associated protein; CDC20, cell division cycle protein 20; KIF20A, kinesin family member 20A; ENO1, enolase 1; P4HA, prolyl 4-hydroxylase subunit α; PRDX1, peroxiredoxin 1; FAM162A, family with sequence similarity 162, member A; STAT1, signal transducer and activator of transcription 1; CDK1, cyclin-dependent kinase 1; TACC3, transforming, acidic coiled-coil containing protein 3; PRKDC, protein kinase, DNA-activated, catalytic polypeptide; H2AFY, H2A histone family, member Y; SLC16A1, solute carrier family 16, member 1; VEGF, vascular endothelial growth factor; HIF, hypoxia inducible factor; PLOD2, procollagen-lysine,2-oxoglutarate 5-dioxygenase 2; NF-κB, nuclear factor-kappa B.
doi:10.1371/journal.pone.0106801.g008

the basis of pairwise comparisons of histological subtypes, as shown in Fig. 4.

Survival analysis in UPS patients

We used the logrank test to analyze the survival of UPS patients. We selected the best p value for various thresholds (30th, 40th, 50th, 60th, 70th, and 80th percentiles) of gene expression signals in UPS patients for each probe set when the gene expression signals were binarized. Adjusted p values were obtained by adjusting the data for the multiple testing problem (6 thresholds $\times$29 probe sets) based on the permutation test, as shown in Table S6. Only *STAT1* showed a statistically significant association with survival in UPS (logrank p value 1.84×10^{-6} and adjusted p value 2.99×10^{-3} after the permutation test). Fig. 7 shows that *STAT1*-positive and *STAT1*-negative groups had clearly different survival curves based on the Kaplan-Meier method.

Discussion

In the present study, we conducted a simulation based on a permutation test to extract genes that are both diagnostic markers (for discrimination of histological subtypes) and prognostic markers (for overall survival in STS). As shown in Table 2, 25 genes were extracted, and their adjusted p values were statistically significant (adjusted $p<0.05$). We analyzed studies related to these 25 genes and found many reports suggesting that these 25 genes are effective prognostic/predictive factors or therapeutic targets, as shown in Table S7, according to the literature (See Supplementary Discussion).

Although we did not try to identify the molecular mechanisms behind the 25 selected genes, several published studies have examined pathways related to these 25 genes, as shown in Table S7 and Fig. 8. These 25 genes are roughly classified into 4 types, namely, hypoxia-related genes (*MIF*, *SCD1*, *P4HA1*, *ENO1/MBP1*, *FAM162A/HGTD-P*, *SLC16A1/MCT1*, *FN1*, and *STAT1*), cell cycle- and DNA repair-related genes (*ASPM*, *CDK1/CDC2*, *CDC20*, *KIF20A/MKlp2*, *PTTG1*, *TACC3*, *PRDX1*, *PRKDC/DNA-PKcs*, and *H2AFY/H2AX*), growth factor signal transduction-related genes, and other genes. Cell cycle- and DNA repair-related genes, hypoxia-induced genes, and growth factor signal transduction-related genes are key players in tumor growth, angiogenesis, metabolism, invasion, and metastasis in various types of cancer. In fact, these processes are attenuated by the inhibition or silencing of many of these 25 genes, as shown in Table S7. These genes are therefore possible prognostic/predictive markers and/or therapeutic targets.

STAT1 expression was found to be strongly associated with survival in UPS patients. STAT1 interacts directly with p53 and induces cell growth arrest and apoptosis, as shown in Fig. 8. Although *STAT1* is repressed by HIF-1, the *STAT1*-positive group among the UPS patients had a better prognosis, even when hypoxia-related genes were upregulated. Therefore, *STAT1* is a possible novel, independent prognostic/predictive factor of STS, particularly UPS.

In the diagnosis of STS, classification of UPS is the most controversial topic. Among the 25 selected genes, hypoxia-related genes (*MIF*, *SCD1*, *P4HA1*, *ENO1/MBP1*, *FAM162A/HGTD-P*, *SLC16A1/MCT1*, *FN1*, and *STAT1*) are present in this study. In particular, the genes *MIF*, *SCD1*, *P4HA1*, *ENO1/MBP1*, and *FAM162A/HGTD-P* are differentially expressed between UPS and MFS, as shown in Fig. 6 and Table 4. Furthermore, *STAT1* is a prognostic marker in UPS patients, as shown in Fig. 7. Therefore, these hypoxia-related genes are promising prognostic and therapeutic targets and, if validated, may improve the treatment/diagnosis of this type of cancer. Further research is needed regarding the hypoxia-related pathways in highly malignant STS.

We manually constructed a hypothetical regulation model (Figure 8) of metabolic and signaling control in highly malignant STS. Nevertheless, according to the literature, a part of these networks could be automatically predicted by pathway and interaction analyses. For example, pathways of the cell cycle and the DNA damage response were identified by IntPath [33,60,61] with statistical significance (q value <0.05), as shown in Table S8. Interaction networks of the cell cycle (*ASPM*, *CDK1*, *CDC20*, *KIF20A*, *PTTG1*, *PRKDC*, and *TACC3*) and *HIF-1* (*MIF*, *ENO1*, and *PRDX1*) were identified by means of STRING [62], as shown in Fig. S1. Nonetheless, these tools should be used with appropriate parameters [34,60,61]. Such tools are more effective methods when large numbers of candidate genes are extracted.

In summary, we analyzed microarray gene expression data from 88 STS patients using a combination method involving knowledge-based filtering and a simulation based on the integration of multiple statistics to reduce multiple testing problems. Our combination method automatically identified 25 genes in the gene expression data from STS. These genes showed significant differential expression between different histological subtypes, including UPS, and showed associations with survival in STS. Furthermore, we conducted a bibliographic survey in terms of cancer progression for the 25 identified genes, and substantial evidence was uncovered in the literature. These genes were roughly classified into 4 types, namely, hypoxia-related genes, cell cycle- and DNA repair-related genes, growth factor signal transduction-related genes, and other genes. *STAT1* showed a statistically significant association with the survival of UPS patients (logrank adjusted $p = 0.00299$). Although only a few studies have investigated the association of these genes with survival in STS, many recent studies have reported that these genes are prognostic factors and/or therapeutic targets in other types of cancers. Therefore, these results suggest that our combination method is capable of identifying genes that are potential prognostic/predictive factors and/or therapeutic targets in STS and possibly in other cancers. These disease-associated genes deserve further preclinical and clinical validation.

Supporting Information

Figure S1 The pathways predicted by STRING from the 25 selected genes.
(PDF)

Table S1 Clinical data of the 88 patients with soft tissue sarcoma. UPS: undifferentiated pleomorphic sarcoma, MLS: myxoid liposarcoma, SS: synovial sarcoma, MFS: myxofibrosarcoma, LMS: leiomyosarcoma, FS: fibrosarcoma, MPNST: malignant peripheral nerve sheath tumor, Tumor metastasis indicates the incidence of tumor metastasis in STS patients.
(XLS)

Table S2 The MIM number list.
(XLS)

Table S3 Selected Affymetrix probe IDs.
(XLS)

Table S4 Information on PCA, including the eigenvector, standard deviation, proportion of variance, and cumulative proportion for 29 probe sets. PCA: principal component analysis, PC: principal components.
(XLS)

Table S5 Information on PCA, including the eigenvector, standard deviation, proportion of variance, and cumulative proportion for 9 probe sets. PCA: principal component analysis, PC: principal components.
(XLS)

Table S6 Survival analysis in UPS using the logrank test. Adjusted p values were calculated using the permutation test (100,000 repeats) from logrank p values.
(XLS)

Table S7 Gene or pathway annotations and likelihood as prognostic/predictive factors and/or therapeutic targets. Adjusted p values were calculated using the permutation test (100,000 repeats) from logrank p values.
(XLS)

Table S8 Pathway analysis in IntPath. k: genes from the overlap between genes in the list and genes in the pathway, n: the number of genes in the input gene list, m: the number of genes in the identified pathways, N: the total number of genes. The p values were calculated using the hypergeometric test; the q values were calculated from the p values using the Benjamini-Hochberg (BH) method.
(XLS)

Information S1
(PDF)

Acknowledgments

The authors thank Professor Yasunori Machida (Nagoya University, Japan) and the Laboratory Head Hitoshi Ichikawa (National Cancer Center Research Institute, Japan) for the helpful discussions.

Author Contributions

Conceived and designed the experiments: HT NI HS TY TH. Performed the experiments: RN SS RI AD YI TT SM KY TN. Analyzed the data: AT HT. Contributed reagents/materials/analysis tools: AT RN HT TH. Contributed to the writing of the manuscript: AT HT.

References

1. Trojani M, Contesso G, Coindre JM, Rouesse J, Bui NB, et al. (1984) Soft-tissue sarcomas of adults; study of pathological prognostic variables and definition of a histopathological grading system. Int J Cancer 33: 37–42.
2. Guillou L, Coindre JM, Bonichon F, Nguyen BB, Terrier P, et al. (1997) Comparative study of the National Cancer Institute and French Federation of Cancer Centers Sarcoma Group grading systems in a population of 410 adult patients with soft tissue sarcoma. J Clin Oncol 15: 350–362.
3. Clark J, Rocques PJ, Crew AJ, Gill S, Shipley J, et al. (1994) Identification of novel genes, SYT and SSX, involved in the t(X;18)(p11.2;q11.2) translocation found in human synovial sarcoma. Nat Genet 7: 502–508.
4. Antonescu CR, Elahi A, Humphrey M, Lui MY, Healey JH, et al. (2000) Specificity of TLS-CHOP rearrangement for classic myxoid/round cell liposarcoma: absence in predominantly myxoid well-differentiated liposarcomas. J Mol Diagn 2: 132–138.
5. Kohno T, Ichikawa H, Totoki Y, Yasuda K, Hiramoto M, et al. (2012) KIF5B-RET fusions in lung adenocarcinoma. Nat Med 18: 375–377.
6. Lux ML, Rubin BP, Biase TL, Chen CJ, Maclure T, et al. (2000) KIT extracellular and kinase domain mutations in gastrointestinal stromal tumors. Am J Pathol 156: 791–795.
7. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, et al. (2008) Somatic mutations affect key pathways in lung adenocarcinoma. Nature 455: 1069–1075.
8. Helman LJ, Meltzer P (2003) Mechanisms of sarcoma development. Nat Rev Cancer 3: 685–694.
9. Hasegawa T, Yamamoto S, Nojima T, Hirose T, Nikaido T, et al. (2002) Validity and reproducibility of histologic diagnosis and grading for adult soft-tissue sarcomas. Hum Pathol 33: 111–115.
10. Fletcher CD (1992) Pleomorphic malignant fibrous histiocytoma: fact or fiction? A critical reappraisal based on 159 tumors diagnosed as pleomorphic sarcoma. Am J Surg Pathol 16: 213–228.
11. Hollowood K, Fletcher CD (1995) Malignant fibrous histiocytoma: morphologic pattern or pathologic entity? Semin Diagn Pathol 12: 210–220.
12. Fletcher CD, Gustafson P, Rydholm A, Willen H, Akerman M (2001) Clinicopathologic re-evaluation of 100 malignant fibrous histiocytomas: prognostic relevance of subclassification. J Clin Oncol 19: 3045–3050.
13. Nakayama R, Nemoto T, Takahashi H, Ohta T, Kawai A, et al. (2007) Gene expression analysis of soft tissue sarcomas: characterization and reclassification of malignant fibrous histiocytoma. Mod Pathol 20: 749–759.
14. Takahashi H, Nemoto T, Yoshida T, Honda H, Hasegawa T (2006) Cancer diagnosis marker extraction for soft tissue sarcomas based on gene expression profiling data by using projective adaptive resonance theory (PART) filtering method. BMC Bioinformatics 7: 399.
15. Fletcher CDM, Unni KK, Mertens F, editors (2002) Pathology and Genetics of Tumors of Soft Tissue and Bone. Lyon: IARC Press.
16. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
17. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.
18. Takahashi H, Takahashi A, Naito S, Onouchi H (2012) BAIUCAS: a novel BLAST-based algorithm for the identification of upstream open reading frames with conserved amino acid sequences and its application to the *Arabidopsis thaliana* genome. Bioinformatics 28: 2231–2241.
19. Takahashi H, Nakagawa A, Kojima S, Takahashi A, Cha BY, et al. (2012) Discovery of novel rules for G-quadruplex-forming sequences in plants by using bioinformatics methods. J Biosci Bioeng 114: 570–575.
20. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95: 14863–14868.
21. Arima C, Hakamada K, Okamoto M, Hanai T (2008) Modified fuzzy gap statistic for estimating preferable number of clusters in fuzzy k-means clustering. J Biosci Bioeng 105: 273–281.
22. Tomida S, Hanai T, Honda H, Kobayashi T (2002) Analysis of expression profile using fuzzy adaptive resonance theory. Bioinformatics 18: 1073–1083.
23. Takahashi H, Tomida S, Kobayashi T, Honda H (2003) Inference of common genetic network using fuzzy adaptive resonance theory associated matrix method. J Biosci Bioeng 96: 154–160.
24. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102: 15545–15550.
25. Takahashi H, Honda H (2006) Modified signal-to-noise: a new simple and practical gene filtering approach based on the concept of projective adaptive resonance theory (PART) filtering method. Bioinformatics 22: 1662–1664.
26. Takahashi H, Kobayashi T, Honda H (2005) Construction of robust prognostic predictors by using projective adaptive resonance theory as a gene filtering method. Bioinformatics 21: 179–186.
27. Kawamura T, Takahashi H, Honda H (2008) Proposal of new gene filtering method, BagPART, for gene expression analysis with small sample. J Biosci Bioeng 105: 81–84.
28. Ando T, Suguro M, Hanai T, Kobayashi T, Honda H, et al. (2002) Fuzzy neural network applied to gene expression profiling for predicting the prognosis of diffuse large B-cell lymphoma. Jpn J Cancer Res 93: 1207–1212.
29. Takahashi H, Masuda K, Ando T, Kobayashi T, Honda H (2004) Prognostic predictor with multiple fuzzy neural models using expression profiles from DNA microarray for metastases of breast cancer. J Biosci Bioeng 98: 193–199.
30. Takahashi H, Honda H (2005) A new reliable cancer diagnosis method using boosted fuzzy classifier with a SWEEP operator method. J Chem Eng Jpn 38: 763–773.
31. Takahashi H, Murase Y, Kobayashi T, Honda H (2007) New cancer diagnosis modeling using boosting and projective adaptive resonance theory with improved reliable index. Biochem Eng J 33: 100–109.
32. Takahashi H, Honda H (2006) Prediction of peptide binding to major histocompatibility complex class II molecules through use of boosted fuzzy classifier with SWEEP operator method. J Biosci Bioeng 101: 137–141.
33. Zhou H, Jin J, Zhang H, Yi B, Wozniak M, et al. (2012) IntPath–an integrated pathway gene relationship database for model organisms and important pathogens. BMC Syst Biol 6 Suppl 2: S2.
34. Zhou H, Wong L (2011) Comparative analysis and assessment of M. tuberculosis H37Rv protein-protein interaction datasets. BMC Genomics 12 Suppl 3: S20.
35. Kotooka N, Komatsu A, Takahashi H, Nonaka M, Kawaguchi C, et al. (2013) Predictive value of high-molecular weight adiponectin in subjects with a higher risk of the development of metabolic syndrome: From a population based 5-year follow-up data. Int J Cardiol 167: 1068–1070.
36. Takahashi H, Honda H (2006) Lymphoma prognostication from expression profiling using a combination method of boosting and projective adaptive resonance theory. J Chem Eng Jpn 39: 767–771.
37. Takahashi H, Aoyagi K, Nakanishi Y, Sasaki H, Yoshida T, et al. (2006) Classification of intramural metastases and lymph node metastases of esophageal

cancer from gene expression based on boosting and projective adaptive resonance theory. J Biosci Bioeng 102: 46–52.
38. Matsuo N, Mase H, Makino M, Takahashi H, Banno H (2009) Identification of ENHANCER OF SHOOT REGENERATION 1-upregulated genes during in vitro shoot regeneration. Plant Biotechnol 26: 385–393.
39. Yajima I, Kumasaka MY, Naito Y, Yoshikawa T, Takahashi H, et al. (2012) Reduced GNG2 expression levels in mouse malignant melanomas and human melanoma cell lines. Am J Cancer Res 2: 322–329.
40. Sano M, Aoyagi K, Takahashi H, Kawamura T, Mabuchi T, et al. (2010) Forkhead box A1 transcriptional pathway in KRT7-expressing esophageal squamous cell carcinomas with extensive lymph node metastasis. Int J Oncol 36: 321–330.
41. Chiba Y, Mineta K, Hirai MY, Suzuki Y, Kanaya S, et al. (2013) Changes in mRNA stability associated with cold stress in Arabidopsis cells. Plant Cell Physiol 54: 180–194.
42. Nakagawa A, Takahashi H, Kojima S, Sato N, Ohga K, et al. (2012) Berberine enhances defects in the establishment of leaf polarity in asymmetric leaves1 and asymmetric leaves2 of Arabidopsis thaliana. Plant Mol Biol 79: 569–581.
43. Yoshimura K, Mori T, Yokoyama K, Koike Y, Tanabe N, et al. (2011) Identification of alternative splicing events regulated by an Arabidopsis serine/arginine-like protein, atSR45a, in response to high-light stress using a tiling array. Plant Cell Physiol 52: 1786–1805.
44. Portal D, Zhou H, Zhao B, Kharchenko PV, Lowry E, et al. (2013) Epstein-Barr virus nuclear antigen leader protein localizes to promoters and enhancers with cell transcription factors and EBNA2. Proc Natl Acad Sci USA 110: 18537–18542.
45. Takahashi H, Iwakawa H, Nakao S, Ojio T, Morishita R, et al. (2008) Knowledge-based fuzzy adaptive resonance theory and its application to the analysis of gene expression in plants. J Biosci Bioeng 106: 587–593.
46. Takahashi H, Kaniwa N, Saito Y, Sai K, Hamaguchi T, et al. (2013) Identification of a candidate single-nucleotide polymorphism related to chemotherapeutic response through a combination of knowledge-based algorithm and hypothesis-free genomic data. J Biosci Bioeng 116: 768–773.
47. Takahashi H, Sai K, Saito Y, Kaniwa N, Matsumura Y, et al. (2014) Application of a combination of a knowledge-based algorithm and 2-stage screening to hypothesis-free genomic data on irinotecan-treated patients for identification of a candidate single nucleotide polymorphism related to an adverse effect. PLoS One (DOI: 10.1371/journal.pone.0105160)
48. Takahashi H, Iwakawa H, Ishibashi N, Kojima S, Matsumura Y, et al. (2013) Meta-analyses of microarrays of arabidopsis asymmetric leaves1 (as1), as2 and their modifying mutants reveal a critical role for the ETT pathway in stabilization of adaxial-abaxial patterning and cell division during leaf development. Plant Cell Physiol 54: 418–431.
49. Kojima S, Iwasaki M, Takahashi H, Imai T, Matsumura Y, et al. (2011) Asymmetric leaves2 and Elongator, a histone acetyltransferase complex, mediate the establishment of polarity in leaves of *Arabidopsis thaliana*. Plant Cell Physiol 52: 1259–1273.
50. Iwasaki M, Takahashi H, Iwakawa H, Nakagawa A, Ishikawa T, et al. (2013) Dual regulation of ETTIN (ARF3) gene expression by AS1–AS2, which maintains the DNA methylation level, is involved in stabilization of leaf adaxial-abaxial partitioning in Arabidopsis. Development 140: 1958–1969.
51. Takahashi H, Nakayama R, Hayashi S, Nemoto T, Murase Y, et al. (2013) Macrophage migration inhibitory factor and stearoyl-CoA desaturase 1: potential prognostic markers for soft tissue sarcomas based on bioinformatics analyses. PLoS One 8: e78250.
52. Yoshida T, Ono H, Kuchiba A, Saeki N, Sakamoto H (2010) Genome-wide germline analyses on cancer susceptibility and GeMDBJ database: Gastric cancer as an example. Cancer Sci 101: 1582–1589.
53. Mantel N (1966) Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemoth Rep 50: 163–170.
54. Hasegawa T, Yokoyama R, Lee YH, Shimoda T, Beppu Y, et al. (2000) Prognostic relevance of a histological grading system using MIB-1 for adult soft-tissue sarcoma. Oncology 58: 66–74.
55. Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. J Amer Statist Assn 53: 457–481.
56. Flicek P, Amode MR, Barrell D, Beal K, Billis K, et al. (2014) Ensembl 2014. Nucleic Acids Res 42: D749–D755.
57. Pearson K (1901) On Lines and Planes of Closest Fit to Systems of Points in Space. Philos Mag 2 559–572.
58. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc serB 57: 298–300.
59. Chibon F, Lagarde P, Salas S, Perot G, Brouste V, et al. (2010) Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity. Nat Med 16: 781–787.
60. Zhou H, Gao S, Nguyen NN, Fan M, Jin J, et al. (2014) Stringent homology-based prediction of H. sapiens-M. tuberculosis H37Rv protein-protein interactions. Biol Direct 9: 5.
61. Zhou H, Rezaei J, Hugo W, Gao S, Jin J, et al. (2013) Stringent DDI-based prediction of H. sapiens-M. tuberculosis H37Rv protein-protein interactions. BMC Syst Biol 7 Suppl 6: S6.
62. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, et al. (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41: D808–D815.

OPEN ACCESS Freely available online PLOS | ONE

Application of a Combination of a Knowledge-Based Algorithm and 2-Stage Screening to Hypothesis-Free Genomic Data on Irinotecan-Treated Patients for Identification of a Candidate Single Nucleotide Polymorphism Related to an Adverse Effect

Hiro Takahashi[1,2,3]*, Kimie Sai[4], Yoshiro Saito[4], Nahoko Kaniwa[4], Yasuhiro Matsumura[5], Tetsuya Hamaguchi[6], Yasuhiro Shimada[6], Atsushi Ohtsu[7], Takayuki Yoshino[7], Toshihiko Doi[7], Haruhiro Okuda[4], Risa Ichinohe[3,8], Anna Takahashi[2], Ayano Doi[3,8], Yoko Odaka[3], Misuzu Okuyama[3], Nagahiro Saijo[9¤a], Jun-ichi Sawada[10¤b], Hiromi Sakamoto[3], Teruhiko Yoshida[3]

1 Graduate School of Horticulture, Chiba University, Matsudo, Chiba, Japan, 2 Plant Biology Research Center, Chubu University, Kasugai, Aichi, Japan, 3 Division of Genetics, National Cancer Center Research Institute, Tokyo, Japan, 4 Division of Medicinal Safety Science, National Institute of Health Sciences, Tokyo, Japan, 5 Division of Developmental Therapeutics, Research Center for Innovative Oncology, National Cancer Center Hospital East, Kashiwa, Chiba, Japan, 6 Gastrointestinal Medical Oncology Division, National Cancer Center Hospital, Tokyo, Japan, 7 Department of Gastrointestinal Oncology, National Cancer Center Hospital East, Kashiwa, Chiba, Japan, 8 Faculty of Horticulture, Chiba University, Matsudo, Chiba, Japan, 9 National Cancer Center Hospital East, Kashiwa, Chiba, Japan, 10 Division of Functional Biochemistry and Genomics, National Institute of Health Sciences, Tokyo, Japan

Abstract

Interindividual variation in a drug response among patients is known to cause serious problems in medicine. Genomic information has been proposed as the basis for "personalized" health care. The genome-wide association study (GWAS) is a powerful technique for examining single nucleotide polymorphisms (SNPs) and their relationship with drug response variation; however, when using only GWAS, it often happens that no useful SNPs are identified due to multiple testing problems. Therefore, in a previous study, we proposed a combined method consisting of a knowledge-based algorithm, 2 stages of screening, and a permutation test for identifying SNPs. In the present study, we applied this method to a pharmacogenomics study where 109,365 SNPs were genotyped using Illumina Human-1 BeadChip in 168 cancer patients treated with irinotecan chemotherapy. We identified the SNP rs9351963 in potassium voltage-gated channel subfamily KQT member 5 (*KCNQ5*) as a candidate factor related to incidence of irinotecan-induced diarrhea. The *p* value for rs9351963 was 3.31×10^{-5} in Fisher's exact test and 0.0289 in the permutation test (when multiple testing problems were corrected). Additionally, rs9351963 was clearly superior to the clinical parameters and the model involving rs9351963 showed sensitivity of 77.8% and specificity of 57.6% in the evaluation by means of logistic regression. Recent studies showed that *KCNQ4* and *KCNQ5* genes encode members of the M channel expressed in gastrointestinal smooth muscle and suggested that these genes are associated with irritable bowel syndrome and similar peristalsis diseases. These results suggest that rs9351963 in *KCNQ5* is a possible predictive factor of incidence of diarrhea in cancer patients treated with irinotecan chemotherapy and for selecting chemotherapy regimens, such as irinotecan alone or a combination of irinotecan with a KCNQ5 opener. Nonetheless, clinical importance of rs9351963 should be further elucidated.

Citation: Takahashi H, Sai K, Saito Y, Kaniwa N, Matsumura Y, et al. (2014) Application of a Combination of a Knowledge-Based Algorithm and 2-Stage Screening to Hypothesis-Free Genomic Data on Irinotecan-Treated Patients for Identification of a Candidate Single Nucleotide Polymorphism Related to an Adverse Effect. PLoS ONE 9(8): e105160. doi:10.1371/journal.pone.0105160

Editor: Olga Y. Gorlova, Geisel School of Medicine at Dartmouth College, United States of America

Received March 29, 2014; **Accepted** July 17, 2014; **Published** August 15, 2014

Copyright: © 2014 Takahashi et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported in part by the Ministry of Education, Culture, Sports, Science, and Technology of Japan (MEXT): Grants-in-Aid for Scientific Research for Young Scientists (B) (nos. 21710211 and 24710222 to H.T.) and Grant-in-Aid for Scientific Research on Innovative Areas (no. 26114703 to H.T.). This work was also supported by the Advanced Research for Medical Products Mining Program of the National Institute of Biomedical Innovation (NIBIO ID10-41), the Futaba Electronics Memorial Foundation, the Research Foundation for the Electrotechnology of Chubu, and the Nakajima Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: hiro.takahashi@chiba-u.jp

¤a Current address: Japanese Society of Medical Oncology, Tokyo, Japan
¤b Current address: Pharmaceutical and Medical Devices Agency, Tokyo, Japan

Introduction

Genomic information has been proposed to be utilized as the basis for "personalized" health care. Interindividual variation in a drug response among patients has been well documented to cause serious problems in pharmacotherapy. This variation may be due to multiple factors such as disease phenotypes, genetic and clinical parameters (or environmental factors), and variability in the drug target or allergic response; all of these factors may affect both main and side effects [1,2]. Although some biomarkers [3–9] have been proposed, it is still difficult to determine which group of patients will respond positively, which patients are nonresponders, and which may experience adverse reactions in cases where patients are administered the same medication dose. For effectiveness of personalized medicine in cancer chemotherapy, it is critically important to observe interindividual differences in a drug response and the role of genetic polymorphisms relevant to the drug metabolic pathways and drug response biology in pharmacogenomics [10].

Irinotecan (CPT-11), an anticancer prodrug, is widely used for the treatment of a broad range of carcinomas, such as colorectal, lung, ovarian, and cervical cancers. Unexpected severe diarrhea and neutropenia are prominent adverse effects of irinotecan treatment. The active metabolite SN-38 (7-ethyl-10-hydroxycamptothecin), a topoisomerase I inhibitor, is generated via hydrolysis of the parent compound by carboxylesterases [11], and is subsequently glucuronidated by uridine diphosphate glucuronosyltransferases (UGTs), such as UGT1A1, UGT1A7, or UGT1A9, to form an inactive metabolite, SN-38 glucuronide (SN-38G) [12–14]. Irinotecan is also inactivated by CYP3A4 to produce 7-ethyl-10- [4-N-(5-aminopentanoic acid)-1-piperidino] carbonyloxycamptothecin (APC; a major CYP3A4 product) and 7-ethyl-10-(4-amino-1-piperidino) carbonyloxycamptothecin (NPC; a minor product) [15,16]. Irinotecan and its metabolites are excreted into the bile and urine via the action of ATP-binding cassette (ABC) transporters, such as P-glycoprotein (P-gp/ABCB1), multiple resistance-associated protein 2 (MRP2/ABCC2), and breast cancer resistance protein (BCRP/ABCG2) [17]. Transport of SN-38 from the plasma into the liver is mediated by the organic anion transporting polypeptide C (OATP-C/SLCO1B1) [18]. Most of the previous pharmacogenetic studies of irinotecan have been focused on *UGT1A1* polymorphisms and have shown clinical relevance of *UGT1A1*28*, a repeat polymorphism in the TATA box [-54_-39A$(TA)_6$TAA>A$(TA)_7$TAA or -40_-39ins TA], to severe adverse effects [3,19,20]. Based on these findings, in 2005, the Food and Drug Administration (FDA) of the United States approved an amendment for the formulation called Camptosar (irinotecan-HCl) (NDA 20-571/S-024/S-027/S-028) and for clinical use of a genetic diagnostic kit for the **28* allele. In parallel with this advance in the USA, clinical relevance to severe neutropenia of *UGT1A1*6* [211G>A(G71R)], another low-activity allele detected specifically in East Asians, as well as **28*, was demonstrated in several studies on Asian patients [5,21–23]. Accordingly, in June 2008, the Ministry of Health, Labor, and Welfare of Japan approved changes to irinotecan formulations (Campto and Topotecin) by adding a warning about the risk of severe adverse effects in patients either homozygous or compound-heterozygous for *UGT1A1*28* and **6* (**28/*28*, **6/*6*, **28/*6*) and also approved clinical use of a diagnostic kit for *UGT1A1*28* and **6*. Severe adverse effects, however, are reported in patients without the genetic variations **6/*6*, **28/*28*, and **28/*6*; therefore, several clinical studies have suggested that polymorphisms of the drug transporter genes, such as *ABCB1*, *ABCC2*, *ABCG2*, and *SLCO1B1*, might affect irinotecan pharmacokinetics (PK)/pharmacodynamics (PD) in Caucasian and Asian patients [22,24–35], as shown in Fig. 1. Nonetheless, the almost all reported results deal with PK in patients and neutropenia induced by irinotecan as an adverse reaction not but with diarrhea. Therefore, other factors responsible for other irinotecan adverse effects, such as diarrhea should be identified.

Diarrhea induced by irinotecan is classified into early- and delayed-onset diarrhea, occurring within 24 hr or ≤24 hr after irinotecan administration, respectively [36]. Irinotecan induces early-onset diarrhea as one of adverse cholinergic effects (acetylcholinelike effects) by inhibiting acetylcholinesterase (AChE) and binding to muscarinic acetylcholine receptors (mAChR) [37,38]. These inhibitory actions are induced by irinotecan, which has an amino group at the C-10 position [37]. Other than that, irinotecan induces delayed-onset diarrhea via rapid deconjugation of SN38G and adsorption of released free SN-38 by β-glucuronidase of intestinal flora [39–41], as shown in Fig. 1. In the present study, we focused on polymorphisms of genes with transporter activity to identify predictive factors of diarrhea induced by irinotecan because there are many genes related to transporter activity in both pathways.

A genome-wide association study (GWAS), also known as a whole-genome association study (WGA study, or WGAS), is an examination of many common genetic variants in different individuals to determine whether a particular variant is associated with a trait. GWAS using hypothesis-free genomic data is a powerful technique for identifying interindividual variation among patients. On the other hand, multiple testing problems are a limitation of this approach. To address this issue, we recently proposed a combined method consisting of a knowledge-based algorithm, 2 stages of screening, and a permutation test for identifying single nucleotide polymorphisms (SNPs) [7]. In general, the objective of a statistical or bioinformatic analysis is the enrichment of important information in a large dataset [42–47]. The use of a knowledge-based algorithm is not a novel concept, but is both practical and useful [48–66]. In the previous study, we found that rs2293347 in the gene of human epidermal growth factor receptor (*EGFR*) is a candidate SNP related to the chemotherapeutic response; we achieved this result by applying our combined method to gastric cancer patients who were treated with fluoropyrimidine [7]. However, our combined method was applied to only 1 dataset. Therefore, the usability of our combined method as a novel approach was still unclear.

We used the combined method in an actual genome-wide pharmacogenomics study of antitumor drugs, particularly irinotecan. We found that rs9351963 in the gene of potassium voltage-gated channel subfamily KQT member 5 (*KCNQ5*) is a candidate SNP related to the adverse response. Rs9351963 may be a potential predictive factor of incidence of diarrhea in cancer patients treated with the cancer prodrug irinotecan.

Materials and Methods

Ethics statement

The study was conducted according to the principles expressed in the Declaration of Helsinki, and the ethics committees of the National Cancer Center and National Institute of Health Sciences, Japan, approved the study protocol. All patients provided written informed consent to participate.

Preparation of hypothesis-free genomic data for cancer patients treated with irinotecan

This study was performed within the framework of the Millennium Genome Project in Japan, and 4 antitumor drugs

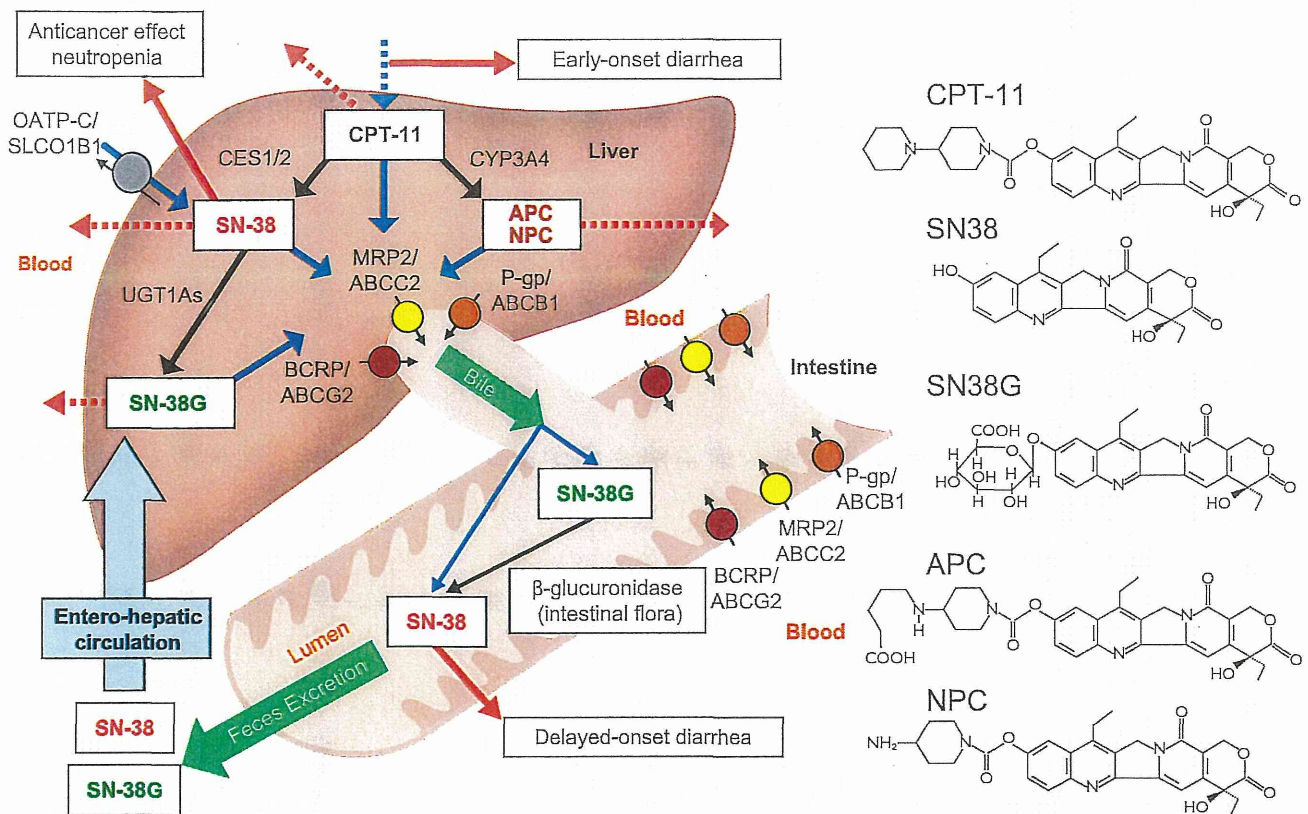


Figure 1. Drug metabolic pathways and the drug response of irinotecan.
doi:10.1371/journal.pone.0105160.g001

were chosen as project targets: gemcitabine, paclitaxel, fluoropyrimidine, and irinotecan. These drugs (alone or in some combination) were administered to approximately 1,000 cancer patients at the National Cancer Center, Japan. Additionally, approximately 1,000 DNA samples were extracted from peripheral blood mononuclear cells and 109,365 SNPs were genotyped using the Illumina Human-1 BeadChip. In this study, we focused on pharmacogenomic properties of irinotecan. Participants included 177 Japanese irinotecan-naïve cancer patients (56 cancer patients treated with irinotecan monotherapy and 121 cancer patients treated with irinotecan combination therapy) at the National Cancer Center Hospital and National Cancer Center Hospital East. A summary of the characteristics of the 176 patients is listed in Table S1. We excluded 1 patient who refused grading of adverse reactions. Furthermore, we excluded 8 patients who did not have genotyping data. Therefore, we analyzed the remaining 168 patients (53 cancer patients treated with irinotecan monotherapy and 115 cancer patients treated with irinotecan combination therapy) in the present study. We defined the 53 patients treated with irinotecan monotherapy as the first dataset and the 168 patients treated with irinotecan chemotherapy (consisting of irinotecan monotherapy and combination therapy) as the second dataset for 2 stages of screening.

Monitoring and adverse effects

A complete medical history and data on physical examination were recorded before the irinotecan therapy. Complete blood cell counts with differentials and platelet counts, as well as blood biochemical variables, were measured once a week during the first 2 months of irinotecan treatment. Adverse events were graded according to the National Cancer Institute - Common Toxicity Criteria (NCI-CTC Version 2.0). Only the highest grade of adverse events was recorded during the first 2 months of irinotecan treatment for each patient and adverse event.

Patient characteristics and clinical parameters

A summary of the patients' characteristics in the two datasets for diarrhea is shown in Table 1. The association of genetic or clinical parameters with incidence of grade ≥2 diarrhea was examined on the basis of Spearman's rank correlation coefficient. "*UGT1A1*6* or **28*" is an effective genetic predictive factor of irinotecan-induced neutropenia and pharmacokinetics in cancer patients [5]. This factor was constructed from 2 polymorphisms: *UGT1A1*6* and **28*.

Fisher's exact test

This statistical test is usually used to determine nonrandom associations between 2 categorical variables [67]. Fisher's exact test is similar to the chi-squared test. If a sample size is large, then the chi-squared test is suitable. Nevertheless, significance values from the chi-squared test are only approximated. Fisher's exact test is used in to analyze contingency tables when the sample sizes are small [67]. We used Fisher's exact test in the present study. The odds ratio (OR) is defined as $a \times d/(b \times c)$, where a is the number of patients that had adverse events with a minor allele, b is the number of patients that did not have adverse events with a minor allele, c is the number of patients that had adverse events with a major allele, and d is the number of patients that did not have adverse events with a major allele. The null hypothesis for Fisher's exact test is OR = 1.

Table 1. Irinotecan-treated cancer patients with SNP information, genetic factor, and clinical parameters for incidence of diarrhea.

Parameters		Diarrhea							
		Irinotecan monotherapy				Irinotecan chemotherapy (including monotherapy)			
		Number of patients		Spearman's rank correlation		Number of patients		Spearman's rank correlation	
		Grade <2	Grade ≥2	ρ	p value	Grade <2	Grade ≥2	ρ	p value
*UGT1A1*6* or **28*	0	15	5	0.056	6.89E–01	64	17	0.009	9.06E–01
	1	21	7			57	16		
	2	3	2			11	3		
Gender	Male	26	11	−0.114	4.15E–01	101	28	−0.012	8.75E–01
	Female	13	3			31	8		
Age		39	14	0.013	9.29E–01	132	36	0.080	3.02E–01
Area		39	14	0.010	9.45E–01	132	36	−0.054	4.88E–01
PS	<2	38	13	0.106	4.50E–01	130	35	0.039	6.15E–01
	≥2	1	1			2	1		
Smoking	0	37	14	−0.119	3.97E–01	111	30	0.008	9.13E–01
	1	2	0			21	6		
Alcohol	0	33	10	0.149	2.88E–01	90	26	−0.036	6.44E–01
	1	6	4			42	10		
Alb	0	18	10	−0.223	1.08E–01	71	24	−0.108	1.62E–01
	1	21	4			60	12		
	2	0	0			1	0		
Hg	0	14	4	0.061	6.65E–01	58	14	0.040	6.05E–01
	1	22	9			67	20		
	2	3	0			6	1		
	3	0	1			0	1		
	4	0	0			1	0		
GOT	0	33	12	−0.014	9.23E–01	108	32	−0.080	3.05E–01
	1	6	2			22	4		
	2	0	0			2	0		
ALP	0	28	8	0.117	4.05E–01	89	23	0.026	7.38E–01
	1	9	6			38	12		
	2	0	0			2	1		
	3	2	0			3	0		
Cr	0	31	13	−0.157	2.62E–01	124	35	−0.060	4.41E–01
	1	8	1			8	1		
C_{max}/dose		39	14	0.049	7.31E–01	132	36	0.019	8.10E–01
AUC ratio		39	14	−0.078	5.81E–01	132	36	−0.109	1.60E–01

Table 1. Cont.

Parameters		Diarrhea								
		Irinotecan monotherapy				Irinotecan chemotherapy (including monotherapy)				
		Number of patients		Spearman's rank correlation		Number of patients		Spearman's rank correlation		
		Grade <2	Grade ≥2	ρ	p value	Grade <2	Grade ≥2	ρ		p value
Concomitant drug - 5-FU	0	39	14	NA	NA	106	28	0.026		7.40E−01
	1	0	0			26	8			
Concomitant drug - CDDP	0	39	14	NA	NA	76	24	−0.076		3.28E−01
	1	0	0			56	12			
Concomitant drug - MMC	0	39	14	NA	NA	121	36	−0.138	†	7.40E−02
	1	0	0			11	0			
Concomitant drug - VP16	0	39	14	NA	NA	129	35	0.014		8.61E−01
	1	0	0			3	1			
Concomitant drug - Amrubicin	0	39	14	NA	NA	132	34	0.210	*	6.25E−03
	1	0	0			0	2			

"*UGT1A1*6* or **28*" is a genetic factor constructed from 2 polymorphisms (UGT1A1**6* and **28*); "2" indicates **6/*6*, **28/*28* or **6/*28*, "1" indicates **6* or **28*, and "0" indicates "other than 2 and 1." Area: body surface area (m^2), PS: performance status, Cr: grade of creatinine, Hg: grade of hemoglobin, Alb: grade of albumin, ALP: grade of alkaline phosphatase, and GOT: grade of glutamic oxaloacetic transaminase. Each laboratory test value (Alb, Hg, GOT, ALP, and Cr) was recorded before the irinotecan therapy. For each type of clinical tests the grade and aberrant values were defined according to the National Cancer Institute - Common Toxicity Criteria (NCI-CTC, Version 2.0). C_{max}/dose: SN38 C_{max}/dose [$10^{-3} \times m^2/L$]. AUC: area under the concentration-time curve. AUC ratio: Ratio of AUC_{SN38}/AUC_{CPT-11}. 5-FU: 5-fluorouracil, CDDP: cisplatin, MMC: mitomycin C, VP16: etoposide. * and † indicate $p<0.05$ and $0.05 \leq p < 0.10$, respectively. For each concomitant drug, 0 means "not administered," 1 indicates administered.

doi:10.1371/journal.pone.0105160.t001

The permutation test

The permutation test theory evolved from the works of Fisher and Pitman in the 1930s [68]. In this study, p values of multiple-comparison analyses were adjusted by applying the permutation test to 2 stages of screening. The case–control (or phenotype) labels were randomly shuffled for the 2 screening stages, and p values were calculated using Fisher's exact test. The lowest p value was selected for the randomized data. This procedure was repeated 100,000 times. Exact p values for the permutation test were calculated based on the distribution of the lowest p values.

Multiple testing correction

The Bonferroni correction is a method used to address the problem of multiple comparisons (also known as the multiple testing problem). It is considered the simplest and most conservative method for control of the family-wise error rate (FWER). In addition, false discovery rate (FDR) controlling procedures, such as the Benjamini-Hochberg (BH) method [69], are more powerful (i.e., less conservative) than the FWER procedures, such as the Bonferroni correction, at the cost of increasing the likelihood of false positives within the rejected hypothesis. In the present study, the BH method was used to calculate the q value. The q value is defined as an FDR analog of the p value.

The Akaike information criterion (AIC)

The AIC is a measure of the relative goodness of fit of a statistical model [70]. A smaller AIC indicates a better fit when comparing fitted objects. The AIC is defined according to the formula $-2\times$ (log likelihood) + $(2\times n_{par})$, where n_{par} represents the number of parameters in the fitted model, and the log-likelihood value [71] is obtained from the logistic regression model.

The receiver operating characteristic (ROC)

ROC analysis is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. It is built by plotting sensitivity (the number of true positive results divided by the number of true positive samples) against (1 minus specificity) at various threshold settings. (Specificity is the number of true negative results divided by the number of true negative samples.) The area under the curve (AUC) of a ROC curve is an indicator of expected performance of the test. A higher AUC is more desirable, with a value of 1.00 denoting perfect performance (sensitivity and specificity are both 100%), while a value of 0.50 indicates random performance.

Gene set based on gene ontology GO terms

GO has been developed to provide scientists with a controlled terminology system for labeling gene functions in a precise, reliable, computer-readable manner. Data for annotated genes and associated GO terms were obtained from the GO website (http://www.geneontology.org). We compiled a GO term list to select polymorphisms in genes with transporter activity (GO:0005215) and related activities, as shown in Table S2. The numbers of GO terms obtained was 943. GO data were obtained on July 1, 2010.

Results

Association analysis of adverse affects and clinical parameters (or a genetic factor)

The association between clinical parameters (or a genetic factor) and incidence of grade $\geq$2 diarrhea was examined on the basis of Spearman's rank correlation coefficient, as shown in Table 1. This table shows that no parameter was associated with the adverse response to chemotherapy (incidence of grade $\geq$2 diarrhea) in the first dataset (*patients treated with irinotecan monotherapy*). Nonetheless, Amrubicin ($p = 0.00625$) was significantly associated with the response in the second dataset (patients treated with any irinotecan chemotherapy: a combination or monotherapy). Mitomycin C (MMC; $p = 0.0740$) was weakly associated with the response. These clinical factors should be evaluated when constructing diagnostic models involving multiple factors.

Extraction of candidate SNPs using the combined method consisting of the knowledge-based algorithm, 2 stages of screening, and the permutation test

In this study, we applied the combined method to hypothesis-free genomic data on cancer patients treated with irinotecan chemotherapy as shown in Fig. 2. Figure 2A shows an outline of the knowledge-based algorithm for identifying SNPs (KB-SNP). In the previous study, we extracted rs numbers (SNP IDs) related to cancer using a combination of National Center for Biotechnology Information (NCBI) dbSNP and NCBI PubMed [7]. In the present study, we extracted rs numbers from genes linked to specific GO terms instead of the combination of NCBI dbSNP and PubMed. In this analysis, we defined specific GO terms as the terms related to transporter activity.

A total of 6,506 SNPs related to transporter activity were extracted from 109,365 SNPs using KB-SNP (Fig. 2B). Furthermore, we excluded SNPs with a p value <0.2 in the Hardy-Weinberg equilibrium (HWE) or the minor allele frequency (MAF) <0.05. Then the extracted 5,242 SNPs were used in the association study.

We analyzed 53 patients treated with irinotecan monotherapy as the first dataset for first-stage screening in the association study (Fig. 2C). Each p value was calculated using Fisher's exact test for the allele model. A total of 24 SNPs with $p<0.005$ were extracted. In the second stage of screening, 168 patients treated with irinotecan chemotherapy (including 53 patients treated with irinotecan monotherapy) were analyzed to validate these 24 SNPs. Adjustment of a calculated p value for the second stage of screening was conducted using the permutation test for these 2 stages of screening (Fig. 2D). Only rs9351963 in *KCNQ5* showed a statistically significant p value (0.0289), which was determined using the permutation test. The rs9351963 is a common variant (MAF = 0.328). Furthermore, we conducted Fisher's exact test and used the Benjamini-Hochberg method [69] to calculate p and q values for the second dataset only. Seven SNPs had a q value <1, as shown in Table 2. Six SNPs (rs11022922, rs3918305, rs3813627, rs768172, rs3813628, and rs10815019) had $q = 0.802$ as shown in Table 2. This result indicates that 5 out of 6 SNPs were false positive; however, we assessed performance of only rs9351963 in the process of model construction.

Comparison of models based on rs9351963 in *KCNQ5*

We analyzed not only an allele model but also dominant and recessive models of rs9351963 in *KCNQ5* in relation to the first dataset (irinotecan monotherapy), the second dataset (any irinotecan chemotherapy), and the dataset of irinotecan combination chemotherapy (excluding irinotecan monotherapy), as shown in Figure 3. Figure 3A shows that the p value of the allele model was the lowest ($p = 8.86\times10^{-5}$, OR = 6.3), and the p value ($p = 1.29\times10^{-4}$, OR = 24) of the dominant model was lower than the p value ($p = 0.0358$, OR = 7.0) of the recessive model in the first dataset. In addition, Figure 3B shows that the p value of the

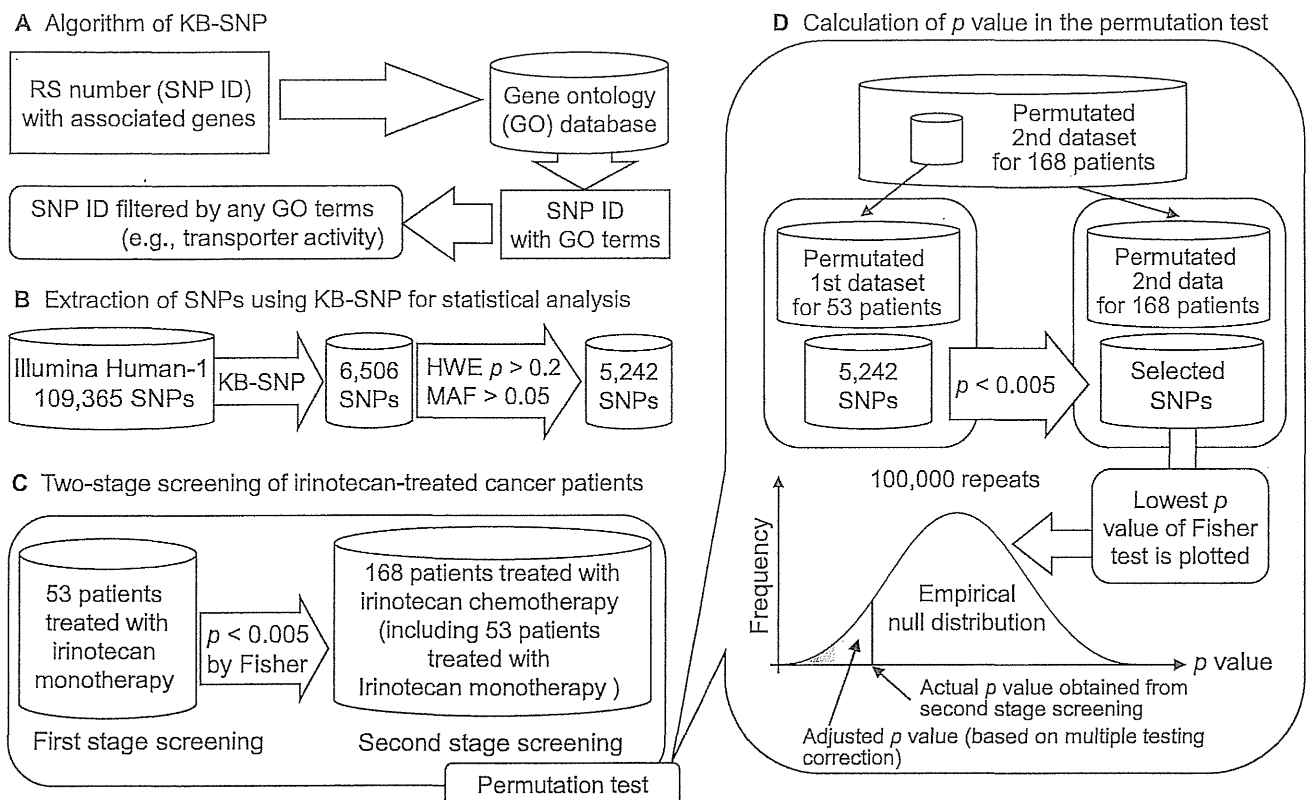


Figure 2. An outline of chemotherapeutic response analysis in irinotecan-treated cancer patients using a combined method consisting of a knowledge-based algorithm for identifying SNPs (KB-SNP), 2 stages of screening, and the permutation test. (A) The KB-SNP algorithm. (B) Extraction of SNPs using KB-SNP for statistical analysis. (C) Two-stage screening of irinotecan-treated cancer patients. (D) Calculation of the p value in the permutation test based on the 2 stages of screening. The second dataset (includes first dataset) was permutated. The permutated first dataset was extracted from the permutated second dataset. By using Fisher's exact test, SNPs with $p<0.005$ for the first dataset were selected from among 5,242 SNPs. Among the selected SNPs, those with the lowest p value in Fisher's exact test for the second dataset were selected. This procedure was repeated 100,000 times and empirical null distribution was constructed. Using the distribution, the actual p value obtained from the second stage of screening was converted to the adjusted p value (based on correction of multiple testing problems). At these screening steps, allele models were used for each SNP.
doi:10.1371/journal.pone.0105160.g002

allele model was the lowest ($p = 3.31\times10^{-5}$, OR = 3.1), and the p value ($p = 1.28\times10^{-4}$, OR = 6.7) of the recessive model was lower than the p value ($p = 4.44\times10^{-3}$, OR = 3.3) of the dominant model in the second dataset. Therefore, we evaluated the 3 models using the dataset of irinotecan combination chemotherapy (excluding irinotecan monotherapy; Fig. 3C). Figure 3C shows that the p value ($p = 1.44\times10^{-3}$, OR = 6.9) of the recessive model meant strong statistical significance and the OR was almost equal to OR (= 7.0) in the first dataset, as shown in Figure 3A. Although ORs of the recessive models seemed to have high homogeneity among all 3 datasets, there was no statistical evidence. Therefore, the proportional odds model was used to construct multiple logistic regression models.

Selection of the model of rs9351963 in *KCNQ5* and construction of multiple regression models

We compared the AICs and AUCs using the second dataset in the 8 models: NULL (without parameter), "*UGT1A1*6* or **28*" (an integrated predictive factor based on polymorphisms related to neutropenia), and rs9351963 (genotype of rs9351963 in *KCNQ5*), Amrubicin, MMC, rs9351963+Amrubicin, rs9351963+MMC, and rs9351963+Amrubicin+MMC (Fig. 4A). Figure 4A shows that performance of all models except *UGT1A1 *6* or **28* is better than the performance of the NULL model. Although the Amrubicin+MMC (combination of Amrubicin and MMC) model was better than Amrubicin alone or MMC, the rs9351963 models were clearly better than the Amrubicin+MMC model, as shown in Figures 4A and 4B. Performance of rs9351963+Amrubicin and rs9351963+MMC models was better than performance of the rs9351963 model. Furthermore, performance of the rs9351963+ Amrubicin+MMC model was better than that of rs9351963+ Amrubicin and rs9351963+MMC models. Therefore, we selected the rs9351963+Amrubicin+MMC model as the best one on the basis of AIC. AUC, sensitivity, and specificity of this model were 0.744, 77.8%, and 57.6% in in the ROC curve, respectively, as shown in Figures 4A and 4B.

Discussion

In the present study, we used 2 stages of screening: the method that is based on the concept of joint analysis. Joint analysis is more efficient than replication-based analysis [72]. The first dataset is a part of the second dataset in joint analysis (the latter includes the former). In contrast, the 2 datasets must be independent in a replication-based analysis (which we did not use here). Our 2 stages of screening derived from the joint analysis were used to increase statistical detection power. KB-SNP was performed prior to 2 stages of screening. KB-SNP reduced the number of

Table 2. Extracted 7 SNPs with $q<1$ for the second dataset.

RS number	Allele	MAF	SNP function		Chr	Position[a]	Associated gene symbol	For second dataset		Two stages of screening	
			Type	Location				p_F	q_{BH}		p_{per}
rs9351963	A/C	0.328	cSNP	intron	6	73749861	*KCNQ5*	3.31E–05	0.173	*	0.0289
rs11022922	C/T	0.376	cSNP	intron	14	63472498	*KCNH5*	3.21E–04	0.802		1.0000
rs3918305	A/G	0.402	cSNP	intron	12	109331162	*SVOP*	6.21E–04	0.802		1.0000
rs3813627	G/T	0.435	cSNP	NearGene-5	1	161195148	*TOMM40L*	7.62E–04	0.802		1.0000
rs768172	A/T	0.441	cSNP	intron	7	95805703	*SLC25A13*	7.87E–04	0.802		1.0000
rs3813628	A/C	0.436	cSNP	5'UTR	1	161196166	*TOMM40L*	1.02E–03	0.802		1.0000
rs10815019	A/G	0.222	cSNP	intron	9	4547288	*SLC1A1*	1.20E–03	0.802		1.0000

RS number: reference SNP identification number in dbSNP, MAF: minor allele frequency, Chr: chromosome number, i.e., a position in human genome GRCh37.p10 build 104, p_F indicates a p value calculated using Fisher's exact test, q_{BH} indicates adjusted p_F value by the Benjamini-Hochberg method, p_{per} indicates p values adjusted using a permutation test for multiple testing problems, * indicates $p_{per}<0.05$. NearGene-5 indicates that the SNP is within 2 kb upstream of a gene.

doi:10.1371/journal.pone.0105160.t002

candidate SNPs to 6,506 from 109,365. Approximately 80,000 SNPs can be extracted without knowledge-based reduction of the SNP number. Thus, statistically significant SNPs cannot be extracted from the present data. We could find the statistically significant rs9351963 in *KCNQ5* by applying the combined method to hypothesis-free genomic data.

The KCNQ/K(v)7 potassium channel family consists of 5 members of neural muscarine channel (M channel; from KCNQ1 to KCNQ5) which have a distinct expression pattern and a functional role. Although KCNQ1 is prevalently expressed in the cardiac muscle, KCNQ2, KCNQ3, KCNQ4, and KCNQ5 are expressed in neural tissue [73–75]. On the other hand, a recent study revealed that KCNQ4 and KCNQ5 are the most abundantly expressed KCNQ channels in smooth muscle throughout the gastrointestinal tract [76]. Furthermore, Jepps et al. opined that drugs that selectively block KCNQ4/KCNQ5 might be promising as therapeutics for the treatment of motility disorders such as constipation associated with irritable bowel syndrome [76]. In other words, drugs that selectively open KCNQ4/KCNQ5 might be effective against diarrhea. The KCNQ family gene products assemble as homomeric or heteromeric tetramers to form functional channels that mediate the M-current [77], a current that is suppressed by mAChR activation [78,79]. Irinotecan induces adverse cholinergic effects (acetylcholinelike actions) by inhibiting AChE and binding to mAChR [37,38,80]. Therefore, polymorphisms of *KCNQ5* genes possibly effect incidence of diarrhea as interindividual variation in the drug response among cancer patients treated with irinotecan chemotherapy.

In the present study, only the highest grade of adverse events is recorded during the first 2 months of irinotecan treatment for each patient and each adverse effect. Therefore, incidence of grade ≥2 diarrhea possibly includes cases caused partially by enterohepatic circulation of APC and NPC, but genotype of rs9351963 in *KCNQ5* correlates with the start date of treatment with antidiarrheal agents (Spearman's rank correlation coefficient $\rho=-0.198$, $p=0.00995$). In other words, genotype of rs9351963 may correlates with the diagnosis (or presentiment) of irinotecan induced early-onset diarrhea (diagnosis is made by trained chemotherapists).

The rs9351963 A>C polymorphism is located in an intron, which does not change the amino acid sequence of the protein and may not influence the biological function of the protein itself. Nonetheless, some intronic polymorphisms are effective markers: For example, rs2237892 in intron 15 of *KCNQ1* is associated with susceptibility to type 2 diabetes mellitus in Japanese individuals [81], and the CA simple sequence repeat in intron 1 (CA-SSR1) of the gene of epidermal growth factor receptor (*EGFR*) is associated with the clinical outcome in gefitinib-treated Japanese patients with non-small cell lung cancer [82]. Furthermore, variations related to intronic or synonymous SNPs possibly affect mRNA stability, translational kinetics, and splicing, resulting in alterations at the protein level, e.g., changes of structure or function [83–89]. Although rs9351963 does not have a known function, this SNP is a possible predictive factor of adverse effects of irinotecan-based chemotherapy and is possibly linked to some functional polymorphisms in *KCNQ5*. Their clinical importance needs to be further elucidated.

In the present study, we extracted rs9351963, which showed a p value (0.0289) obtained using a combination of 2 stages of screening and a permutation test from SNPs selected by KB-SNP. In the second dataset, the p value of Fisher's exact test was 3.31×10^{-5}, and the q value was 0.173 calculated by correction of Benjamini-Hochberg method, as shown in Table 2. This value is

A Irinotecan monotherapy

Allele model

rs9351963	Diarrhea	
	Grade < 2	Grade ≥ 2
A	63	11
C	15	17

OR (95%CI): 6.3 (2.3-19)
$p = 8.86 \times 10^{-5}$

Dominant model

rs9351963	Diarrhea	
	Grade < 2	Grade ≥ 2
AA	26	1
AC+CC	13	13

OR (95%CI): 24 (3.1-1139)
$p = 1.29 \times 10^{-4}$

Recessive model

rs9351963	Diarrhea	
	Grade < 2	Grade ≥ 2
AA+AC	37	10
CC	2	4

OR (95%CI): 7.0 (0.87-89)
$p = 0.0358$

B Any irinotecan chemotherapy (including irinotecan monotherapy)

Allele model

rs9351963	Diarrhea	
	Grade < 2	Grade ≥ 2
A	192	33
C	72	39

OR (95%CI): 3.1 (1.8-5.6)
$p = 3.31 \times 10^{-5}$

Dominant model

rs9351963	Diarrhea	
	Grade < 2	Grade ≥ 2
AA	69	9
AC+CC	63	27

OR (95%CI): 3.3 (1.4-8.5)
$p = 4.44 \times 10^{-3}$

Recessive model

rs9351963	Diarrhea	
	Grade < 2	Grade ≥ 2
AA+AC	123	24
CC	9	12

OR (95%CI): 6.7 (2.3-20)
$p = 1.28 \times 10^{-4}$

C Irinotecan combination chemotherapy (excluding irinotecan monotherapy)

Allele model

rs9351963	Diarrhea	
	Grade < 2	Grade ≥ 2
A	129	22
C	57	22

OR (95%CI): 2.3 (1.1-4.7)
$p = 0.0211$

Dominant model

rs9351963	Diarrhea	
	Grade < 2	Grade ≥ 2
AA	43	8
AC+CC	50	14

OR (95%CI): 1.5 (0.53-4.5)
$p = 0.478$

Recessive model

rs9351963	Diarrhea	
	Grade < 2	Grade ≥ 2
AA+AC	86	14
CC	7	8

OR (95%CI): 6.9 (1.9-26)
$p = 1.44 \times 10^{-3}$

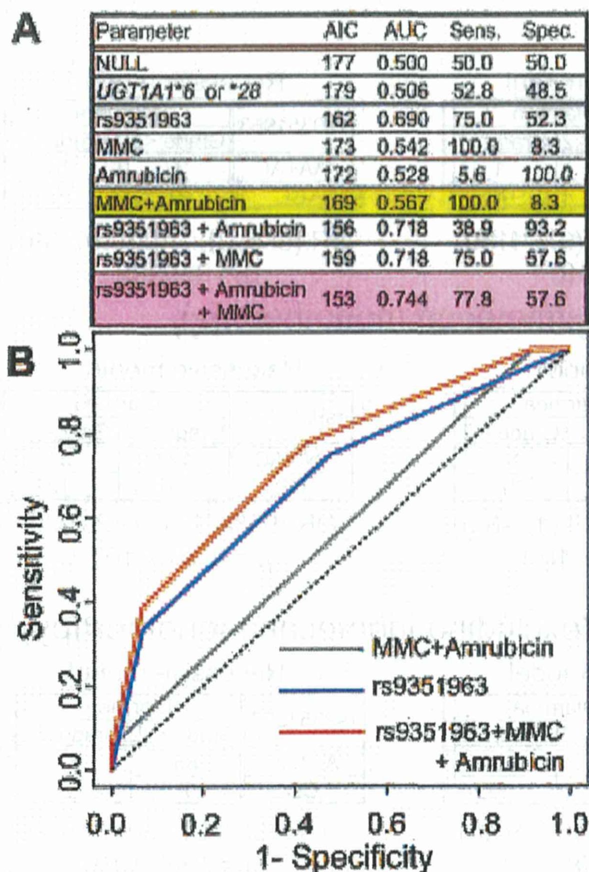
Figure 3. Contingency tables for rs9351963 in *KCNQ5* for each model using each dataset. (A) irinotecan monotherapy (first dataset), (B) any irinotecan chemotherapy (including irinotecan monotherapy; second dataset), and (C) irinotecan combination chemotherapy (excluding irinotecan monotherapy). OR: odds ratio. The *p* values were calculated using Fisher's exact test. CI: confidence interval.
doi:10.1371/journal.pone.0105160.g003

statistically insignificant. Therefore, during the 2 stages of screening, it is statistically sufficient to extract rs9351963.

The calculation of probability of occurrence in Bernoulli trials is suitable to for estimation of validity of the repetition number in the permutation process. In this trial, occurrence probability is defined as ${}_nC_k \times (p_B)^k \times (1 - p_B)^{(n-k)}$, where k is the occurrence number, n is the repetition number, and p_B represents probability. If the repetition number is 100,000 for rs9351963 ($p = 0.02891$ [2891/100000]) and the significance level of the test (α) is 0.05, the occurrence probability is ${}_{100000}C_{2891} \times (0.05)^{2891} \times (1{-}0.05)^{(100000-2891)} = 4.89 \times 10^{-241}$. In statistics, the 99% (or 95%) confidence interval should be considered. The significance level of $\alpha = 0.05$ does not exist in the 99% confidence interval of the p value for rs9351963, because the occurrence probability 4.89×10^{-241} is clearly lower than 0.01. Similarly, if the repetition number is 10,000, the occurrence probability is 3.41×10^{-26}. This way, the occurrence probability is sufficiently low for 10,000 permutations. Nevertheless, we conducted 100,000 permutations to estimate p values more accurately for the permutation test.

Using our combined method involving KB-SNP, we identified rs9351963 as a potential predictive factor of diarrhea in cancer patients treated with irinotecan chemotherapy; however, the comprehensiveness of KB-SNP was limited. Therefore, statistical information regarding the adverse effects of cancer patients treated with irinotecan chemotherapy is shown in Table S3 for incidence of diarrhea ($p < 0.05$) and in Table S4 for incidence of neutropenia ($p < 0.05$). The relevant data are also provided on the website Genome Medicine Database of Japan (GeMDBJ) [90] (https://gemdbj.nibio.go.jp/). These data will be useful for replication studies or meta-analyses in the future.

In conclusion, in the present study, we applied the combined method to hypothesis-free genomic data on cancer patients treated with irinotecan chemotherapy. By means of this method, rs9351963 in *KCNQ5* was extracted as a candidate SNP related to the incidence of diarrhea. For example, the association of rs9351963 with irinotecan-related diarrhea (OR of 3.14) showed a p value of 3.31×10^{-5} in Fisher's exact test (allele model). Even if this p value were adjusted by means of the permutation test for the effects of multiple testing problems, the adjusted p value would still indicate statistical significance (adjusted p value of $0.0289 < 0.05$). Additionally, we evaluated the performance of rs9351963 using multiple regression models. rs9351963 was clearly superior to clinical parameters (or environmental factors) and showed a sensitivity of 77.8% and specificity of 57.6% in the multiple regression model, including rs9351963. Recent studies showed that the *KCNQ4* and *KCNQ5* genes encode components of the M channel expressed in gastrointestinal smooth muscles and suggested that these genes are associated with irritable bowel syndrome and similar peristalsis diseases. These results suggest that rs9351963 may be a predictive factor of diarrhea in cancer patients treated with irinotecan chemotherapy. This SNP may also be useful for selection of chemotherapy regimens, such as irinotecan monotherapy or a combination of irinotecan chemo-



Parameter	AIC	AUC	Sens.	Spec.
NULL	177	0.500	50.0	50.0
*UGT1A1*6* or *28	179	0.506	52.8	48.5
rs9351963	162	0.690	75.0	52.3
MMC	173	0.542	100.0	8.3
Amrubicin	172	0.528	5.6	100.0
MMC+Amrubicin	169	0.567	100.0	8.3
rs9351963 + Amrubicin	156	0.718	38.9	93.2
rs9351963 + MMC	159	0.718	75.0	57.6
rs9351963 + Amrubicin + MMC	153	0.744	77.8	57.6

Figure 4. Comparison of AIC, AUC, and ROC curves for logistic regression models. (A) Parameters of each model. (B) The ROC curve of a model consisting of rs9351963+MMC+ Amrubicin. ROC: receiver operating characteristic, AUC: area under the ROC curve, NULL indicates the model without parameters. Each genetic factor conforms to the proportional odds model, AIC: Akaike's information criterion, AUC: area under the ROC curve, Sens.: Sensitivity (%), Spec.: Specificity (%).
doi:10.1371/journal.pone.0105160.g004

therapy with KCNQ5 opener. Furthermore, the result of the present analysis supports usability of our combined method.

Supporting Information

Table S1 Irinotecan-treated cancer patients, genetic factor, and clinical parameters for incidence of diarrhea and neutropenia. "*UGT1A1*6* or *28*" is a genetic factor constructed from 2 polymorphisms (*UGT1A1*6* and *28*); "2" indicates *6/*6*, *28/*28* or *6/*28*, "1" indicates *6* or *28*, and "0" indicates "other than 2 and 1." Area: body surface area (m^2), PS: performance status, Cr: grade of creatinine, Hg: grade of hemoglobin, Alb: grade of albumin, ALP: grade of alkaline phosphatase, and GOT: grade of glutamic oxaloacetic transaminase. Each laboratory test value (Alb, Hg, GOT, ALP, and Cr) was recorded before the irinotecan therapy. For each type of clinical tests the grade and aberrant values were defined according to the National Cancer Institute - Common Toxicity Criteria (NCI-CTC, Version 2.0). C_{max}/dose: SN38 C_{max}/dose [$10^{-3}\times m^2/L$]. AUC: area under the concentration-time curve. AUC ratio: Ratio of $AUC_{SN38}/AUC_{CPT\text{-}11}$. 5-FU: 5-fluorouracil, CDDP: cisplatin, MMC: mitomycin C, VP16: etoposide. * and † indicate $p<0.05$ and $0.05\leq p<0.10$, respectively. For each concomitant drug, 0 means "not administered," 1 indicates administered.
(XLS)

Table S2 GO term list for transporter activity and the related functions.
(XLS)

Table S3 Statistical information on the chemotherapeutic response (incidence of grade ≥2 diarrhea) of irinotecan-treated cancer patients ($p<0.05$). RS number: reference SNP identification number in dbSNP; *p* values were calculated using Fisher's exact test and *q* values were calculated using the Benjamini-Hochberg (BH) method from *p* values.
(XLS)

Table S4 Statistical information on the chemotherapeutic response (incidence of grade ≥3 neutropenia) of irinotecan-treated cancer patients ($p<0.05$). RS number: reference SNP identification number in dbSNP; *p* values were calculated using Fisher's exact test and *q* values were calculated using the Benjamini-Hochberg (BH) method from *p* values.
(XLS)

Acknowledgments

We thank Ms. Sumiko Ohnami for help with SNP genotyping.

Author Contributions

Conceived and designed the experiments: HT Y. Saito NS JS HS T. Yoshida. Performed the experiments: HT KS NK HO YO MO. Analyzed the data: HT RI AT AD. Contributed reagents/materials/analysis tools: HT YM TH Y. Shimada AO T. Yoshino TD. Contributed to the writing of the manuscript: HT.

References

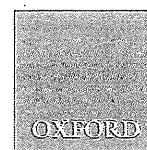
1. Evans WE, McLeod HL (2003) Pharmacogenomics–drug disposition, drug targets, and side effects. N Engl J Med 348: 538–549.
2. Ingelman-Sundberg M (2008) Pharmacogenomic biomarkers for prediction of severe adverse drug reactions. N Engl J Med 358: 637–639.
3. Ando Y, Saka H, Ando M, Sawa T, Muro K, et al. (2000) Polymorphisms of UDP-glucuronosyltransferase gene and irinotecan toxicity: a pharmacogenetic analysis. Cancer Res 60: 6921–6926.
4. Ma F, Sun T, Shi Y, Yu D, Tan W, et al. (2009) Polymorphisms of *EGFR* predict clinical outcome in advanced non-small-cell lung cancer patients treated with Gefitinib. Lung Cancer 66: 114–119.
5. Minami H, Sai K, Saeki M, Saito Y, Ozawa S, et al. (2007) Irinotecan pharmacokinetics/pharmacodynamics and *UGT1A* genetic polymorphisms in Japanese: roles of *UGT1A1*6* and *28*. Pharmacogenet Genomics 17: 497–504.
6. Sato Y, Laird NM, Nagashima K, Kato R, Hamano T, et al. (2009) A new statistical screening approach for finding pharmacokinetics-related genes in genome-wide studies. Pharmacogenomics J 9: 137–146.
7. Takahashi H, Kaniwa N, Saito Y, Sai K, Hamaguchi T, et al. (2013) Identification of a candidate single-nucleotide polymorphism related to chemotherapeutic response through a combination of knowledge-based algorithm and hypothesis-free genomic data. J Biosci Bioeng 116: 768–773.
8. van Kuilenburg AB, Muller EW, Haasjes J, Meinsma R, Zoetekouw L, et al. (2001) Lethal outcome of a patient with a complete dihydropyrimidine dehydrogenase (DPD) deficiency after administration of 5-fluorouracil: frequency of the common IVS14+1G>A mutation causing DPD deficiency. Clin Cancer Res 7: 1149–1153.
9. Raida M, Schwabe W, Hausler P, Van Kuilenburg AB, Van Gennip AH, et al. (2001) Prevalence of a common point mutation in the dihydropyrimidine dehydrogenase (DPD) gene within the 5′-splice donor site of intron 14 in patients with severe 5-fluorouracil (5-FU)- related toxicity compared with controls. Clin Cancer Res 7: 2832–2839.
10. Efferth T, Volm M (2005) Pharmacogenetics for individualized cancer chemotherapy. Pharmacol Ther 107: 155–176.
11. Slatter JG, Su P, Sams JP, Schaaf LJ, Wienkers LC (1997) Bioactivation of the anticancer agent CPT-11 to SN-38 by human hepatic microsomal carboxylesterases and the in vitro assessment of potential drug interactions. Drug Metab Dispos 25: 1157–1164.

12. Iyer L, King CD, Whitington PF, Green MD, Roy SK, et al. (1998) Genetic predisposition to the metabolism of irinotecan (CPT-11). Role of uridine diphosphate glucuronosyltransferase isoform 1A1 in the glucuronidation of its active metabolite (SN-38) in human liver microsomes. J Clin Invest 101: 847–854.
13. Ciotti M, Basu N, Brangi M, Owens IS (1999) Glucuronidation of 7-ethyl-10-hydroxycamptothecin (SN-38) by the human UDP-glucuronosyltransferases encoded at the *UGT1* locus. *Biochem Biophys Res Commun 260*: 199–202.
14. Gagne JF, Montminy V, Belanger P, Journault K, Gaucher G, et al. (2002) Common human *UGT1A* polymorphisms and the altered metabolism of irinotecan active metabolite 7-ethyl-10-hydroxycamptothecin (SN-38). Mol Pharmacol 62: 608–617.
15. Haaz MC, Rivory L, Riche C, Vernillet L, Robert J (1998) Metabolism of irinotecan (CPT-11) by human hepatic microsomes: participation of cytochrome P-450 3A and drug interactions. Cancer Res 58: 468–472.
16. Sai K, Saito Y, Fukushima-Uesaka H, Kurose K, Kaniwa N, et al. (2008) Impact of *CYP3A4* haplotypes on irinotecan pharmacokinetics in Japanese cancer patients. Cancer Chemother Pharmacol 62: 529–537.
17. Sparreboom A, Danesi R, Ando Y, Chan J, Figg WD (2003) Pharmacogenomics of ABC transporters and its role in cancer chemotherapy. Drug Resist Updat 6: 71–84.
18. Nozawa T, Minami H, Sugiura S, Tsuji A, Tamai I (2005) Role of organic anion transporter OATP1B1 (OATP-C) in hepatic uptake of irinotecan and its active metabolite, 7-ethyl-10-hydroxycamptothecin: in vitro evidence and effect of single nucleotide polymorphisms. Drug Metab Dispos 33: 434–439.
19. Iyer L, Das S, Janisch L, Wen M, Ramirez J, et al. (2002) UGT1A1*28 polymorphism as a determinant of irinotecan disposition and toxicity. Pharmacogenomics J 2: 43–47.
20. Innocenti F, Undevia SD, Iyer L, Chen PX, Das S, et al. (2004) Genetic variants in the *UDP-glucuronosyltransferase 1A1* gene predict the risk of severe neutropenia of irinotecan. J Clin Oncol 22: 1382–1388.
21. Han JY, Lim HS, Shin ES, Yoo YK, Park YH, et al. (2006) Comprehensive analysis of *UGT1A* polymorphisms predictive for pharmacokinetics and treatment outcome in patients with non-small-cell lung cancer treated with irinotecan and cisplatin. J Clin Oncol 24: 2237–2244.
22. Jada SR, Lim R, Wong CI, Shu X, Lee SC, et al. (2007) Role of *UGT1A1*6*, *UGT1A1*28* and *ABCG2* c.421C>A polymorphisms in irinotecan-induced neutropenia in Asian cancer patients. Cancer Sci 98: 1461–1467.
23. Sai K, Saito Y, Sakamoto H, Shirao K, Kurose K, et al. (2008) Importance of *UDP-glucuronosyltransferase 1A1*6* for irinotecan toxicities in Japanese cancer patients. Cancer Lett 261: 165–171.
24. Sai K, Saito Y, Maekawa K, Kim SR, Kaniwa N, et al. (2010) Additive effects of drug transporter genetic polymorphisms on irinotecan pharmacokinetics/pharmacodynamics in Japanese cancer patients. Cancer Chemother Pharmacol 66: 95–105.
25. Mathijssen RH, Marsh S, Karlsson MO, Xie R, Baker SD, et al. (2003) Irinotecan pathway genotype analysis to predict pharmacokinetics. Clin Cancer Res 9: 3246–3253.
26. Sai K, Kaniwa N, Itoda M, Saito Y, Hasegawa R, et al. (2003) Haplotype analysis of ABCB1/MDR1 blocks in a Japanese population reveals genotype-dependent renal clearance of irinotecan. Pharmacogenetics 13: 741–757.
27. Zhou Q, Sparreboom A, Tan EH, Cheung YB, Lee A, et al. (2005) Pharmacogenetic profiling across the irinotecan pathway in Asian patients with cancer. Br J Clin Pharmacol 59: 415–424.
28. de Jong FA, Marsh S, Mathijssen RH, King C, Verweij J, et al. (2004) ABCG2 pharmacogenetics: ethnic differences in allele frequency and assessment of influence on irinotecan disposition. Clin Cancer Res 10: 5889–5894.
29. de Jong FA, Scott-Horton TJ, Kroetz DL, McLeod HL, Friberg LE, et al. (2007) Irinotecan-induced diarrhea: functional significance of the polymorphic ABCC2 transporter protein. Clin Pharmacol Ther 81: 42–49.
30. Xiang X, Jada SR, Li HH, Fan L, Tham LS, et al. (2006) Pharmacogenetics of SLCO1B1 gene and the impact of *1b and *15 haplotypes on irinotecan disposition in Asian cancer patients. Pharmacogenet Genomics 16: 683–691.
31. Takane H, Miyata M, Burioka N, Kurai J, Fukuoka Y, et al. (2007) Severe toxicities after irinotecan-based chemotherapy in a patient with lung cancer: a homozygote for the SLCO1B1*15 allele. Ther Drug Monit 29: 666–668.
32. Han JY, Lim HS, Shin ES, Yoo YK, Park YH, et al. (2008) Influence of the organic anion-transporting polypeptide 1B1 (*OATP1B1*) polymorphisms on irinotecan-pharmacokinetics and clinical outcome of patients with advanced non-small cell lung cancer. Lung Cancer 59: 69–75.
33. Han JY, Lim HS, Park YH, Lee SY, Lee JS (2009) Integrated pharmacogenetic prediction of irinotecan pharmacokinetics and toxicity in patients with advanced non-small cell lung cancer. Lung Cancer 63: 115–120.
34. Michael M, Thompson M, Hicks RJ, Mitchell PL, Ellis A, et al. (2006) Relationship of hepatic functional imaging to irinotecan pharmacokinetics and genetic parameters of drug elimination. J Clin Oncol 24: 4228–4235.
35. Sai K, Itoda M, Saito Y, Kurose K, Katori N, et al. (2006) Genetic variations and haplotype structures of the *ABCB1* gene in a Japanese population: an expanded haplotype block covering the distal promoter region, and associated ethnic differences. Ann Hum Genet 70: 605–622.
36. Yang X, Hu Z, Chan SY, Chan E, Goh BC, et al. (2005) Novel agents that potentially inhibit irinotecan-induced diarrhea. Curr Med Chem 12: 1343–1358.
37. Kawato Y, Sekiguchi M, Akahane K, Tsutomi Y, Hirota Y, et al. (1993) Inhibitory activity of camptothecin derivatives against acetylcholinesterase in dogs and their binding activity to acetylcholine receptors in rats. J Pharm Pharmacol 45: 444–448.
38. Hyatt JL, Tsurkan L, Morton CL, Yoon KJ, Harel M, et al. (2005) Inhibition of acetylcholinesterase by the anticancer prodrug CPT-11. Chem Biol Interact 157–158: 247–252.
39. Takakura A, Kurita A, Asahara T, Yokoba M, Yamamoto M, et al. (2012) Rapid deconjugation of SN-38 glucuronide and adsorption of released free SN-38 by intestinal microorganisms in rat. Oncol Lett 3: 520–524.
40. Yamamoto M, Kurita A, Asahara T, Takakura A, Katono K, et al. (2008) Metabolism of irinotecan and its active metabolite SN-38 by intestinal microflora in rats. Oncol Rep 20: 727–730.
41. Kuhn JG (1998) Pharmacology of irinotecan. Oncology (Williston Park) 12: 39–42.
42. Takahashi H, Honda H (2006) Modified signal-to-noise: a new simple and practical gene filtering approach based on the concept of projective adaptive resonance theory (PART) filtering method. Bioinformatics 22: 1662–1664.
43. Takahashi H, Kobayashi T, Honda H (2005) Construction of robust prognostic predictors by using projective adaptive resonance theory as a gene filtering method. Bioinformatics *21: 179–186.*
44. Takahashi H, Iwakawa H, Nakao S, Ojio T, Morishita R, et al. (2008) Knowledge-based fuzzy adaptive resonance theory and its application to the analysis of gene expression in plants. J Biosci Bioeng 106: 587–593.
45. Takahashi H, Honda H (2006) Prediction of peptide binding to major histocompatibility complex class II molecules through use of boosted fuzzy classifier with SWEEP operator method. J Biosci Bioeng 101: 137–141.
46. Kawamura T, Takahashi H, Honda H (2008) Proposal of new gene filtering method, BagPART, for gene expression analysis with small sample. J Biosci Bioeng 105: 81–84.
47. Takahashi H, Takahashi A, Naito S, Onouchi H (2012) BAIUCAS: a novel BLAST-based algorithm for the identification of upstream open reading frames with conserved amino acid sequences and its application to the *Arabidopsis thaliana* genome. Bioinformatics 28: 2231–2241.
48. Chiba Y, Mineta K, Hirai MY, Suzuki Y, Kanaya S, et al. (2013) Changes in mRNA stability associated with cold stress in Arabidopsis cells. Plant Cell Physiol 54: 180–194.
49. Iwasaki M, Takahashi H, Iwakawa H, Nakagawa A, Ishikawa T, et al. (2013) Dual regulation of *ETTIN* (*ARF3*) gene expression by AS1-AS2, which maintains the DNA methylation level, is involved in stabilization of leaf adaxial-abaxial partitioning in *Arabidopsis*. Development 140: 1958–1969.
50. Kojima S, Iwasaki M, Takahashi H, Imai T, Matsumura Y, et al. (2011) ASYMMETRIC LEAVES2 and Elongator, a histone acetyltransferase complex, mediate the establishment of polarity in leaves of *Arabidopsis thaliana*. Plant Cell Physiol 52: 1259–1273.
51. Kotooka N, Komatsu A, Takahashi H, Nonaka M, Kawaguchi C, et al. (2013) Predictive value of high-molecular weight adiponectin in subjects with a higher risk of the development of metabolic syndrome: From a population based 5-year follow-up data. Int J Cardiol 167: 1068–1070.
52. Matsuo N, Mase H, Makino M, Takahashi H, Banno H (2009) Identification of ENHANCER OF SHOOT REGENERATION 1-upregulated genes during in vitro shoot regeneration. Plant Biotechnol 26: 385–393.
53. Nakagawa A, Takahashi H, Kojima S, Sato N, Ohga K, et al. (2012) Berberine enhances defects in the establishment of leaf polarity in *asymmetric leaves1* and *asymmetric leaves2* of *Arabidopsis thaliana*. Plant Mol Biol 79: 569–581.
54. Nakayama R, Nemoto T, Takahashi H, Ohta T, Kawai A, et al. (2007) Gene expression analysis of soft tissue sarcomas: characterization and reclassification of malignant fibrous histiocytoma. Mod Pathol 20: 749–759.
55. Sano M, Aoyagi K, Takahashi H, Kawamura T, Mabuchi T, et al. (2010) Forkhead box A1 transcriptional pathway in KRT7-expressing esophageal squamous cell carcinomas with extensive lymph node metastasis. Int J Oncol 36: 321–330.
56. Yajima I, Kumasaka MY, Naito Y, Yoshikawa T, Takahashi H, et al. (2012) Reduced GNG2 expression levels in mouse malignant melanomas and human melanoma cell lines. Am J Cancer Res 2: 322–329.
57. Yoshimura K, Mori T, Yokoyama K, Koike Y, Tanabe N, et al. (2011) Identification of alternative splicing events regulated by an Arabidopsis serine/arginine-like protein, atSR45a, in response to high-light stress using a tiling array. Plant Cell Physiol 52: 1786–1805.
58. Takahashi H, Iwakawa H, Ishibashi N, Kojima S, Matsumura Y, et al. (2013) Meta-analyses of microarrays of arabidopsis asymmetric leaves1 (as1), as2 and their modifying mutants reveal a critical role for the ETT pathway in stabilization of adaxial-abaxial patterning and cell division during leaf development. Plant Cell Physiol 54: 418–431.
59. Takahashi H, Murase Y, Kobayashi T, Honda H (2007) New cancer diagnosis modeling using boosting and projective adaptive resonance theory with improved reliable index. Biochem Eng J 33: 100–109.
60. Takahashi H, Nemoto T, Yoshida T, Honda H, Hasegawa T (2006) Cancer diagnosis marker extraction for soft tissue sarcomas based on gene expression profiling data by using projective adaptive resonance theory (PART) filtering method. BMC Bioinformatics 7: 399.
61. Takahashi H, Aoyagi K, Nakanishi Y, Sasaki H, Yoshida T, et al. (2006) Classification of intramural metastases and lymph node metastases of esophageal

cancer from gene expression based on boosting and projective adaptive resonance theory. J Biosci Bioeng 102: 46–52.
62. Takahashi H, Honda H (2006) Lymphoma prognostication from expression profiling using a combination method of boosting and projective adaptive resonance theory. J Chem Eng Jpn 39: 767–771.
63. Takahashi H, Honda H (2005) A new reliable cancer diagnosis method using boosted fuzzy classifier with a SWEEP operator method. J Chem Eng Jpn 38: 763–773.
64. Takahashi H, Masuda K, Ando T, Kobayashi T, Honda H (2004) Prognostic predictor with multiple fuzzy neural models using expression profiles from DNA microarray for metastases of breast cancer. J Biosci Bioeng 98: 193–199.
65. Takahashi H, Tomida S, Kobayashi T, Honda H (2003) Inference of common genetic network using fuzzy adaptive resonance theory associated matrix method. J Biosci Bioeng 96: 154–160.
66. Takahashi H, Nakayama R, Hayashi S, Nemoto T, Murase Y, et al. (2013) Macrophage migration inhibitory factor and stearoyl-CoA desaturase 1: potential prognostic markers for soft tissue sarcomas based on bioinformatics analyses. PLoS One 8: e78250.
67. Fisher RA (1922) On the interpretation of $\chi2$ from contingency tables, and the calculation of P. J Roy Statistical Society 85: 87–94.
68. Pitman EJG (1938) Significance tests which may be applied to samples from any population. Part III. The analysis of variance test. Biometrika 29: 322–335.
69. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc serB 57: 298–300.
70. Akaike H (1974) A new look at the statistical model identification. IEEE T Automat Contr 19: 716–723.
71. Sakamoto Y, Ishiguro M, Kitagawa G (1986) Akaike Information Criterion Statistics. Dordrecht: Reidel Publishing Company.
72. Skol AD, Scott LJ, Abecasis GR, Boehnke M (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet 38: 209–213.
73. Delmas P, Brown DA (2005) Pathways modulating neural KCNQ/M (Kv7) potassium channels. Nat Rev Neurosci 6: 850–862.
74. Miceli F, Soldovieri MV, Martire M, Taglialatela M (2008) Molecular pharmacology and therapeutic potential of neuronal Kv7-modulating drugs. Curr Opin Pharmacol 8: 65–74.
75. Brown DA, Passmore GM (2009) Neural *KCNQ* (Kv7) channels. Br J Pharmacol 156: 1185–1195.
76. Jepps TA, Greenwood IA, Moffatt JD, Sanders KM, Ohya S (2009) Molecular and functional characterization of Kv7 K+ channel in murine gastrointestinal smooth muscles. Am J Physiol Gastrointest Liver Physiol 297: G107–115.
77. Schwake M, Jentsch TJ, Friedrich T (2003) A carboxy-terminal domain determines the subunit specificity of KCNQ K+ channel assembly. EMBO Rep 4: 76–81.
78. Cavaliere S, Malik BR, Hodge JJ (2013) KCNQ channels regulate age-related memory impairment. PLoS One 8: e62445.
79. Perez C, Vega R, Soto E (2010) Phospholipase C-mediated inhibition of the M-potassium current by muscarinic-receptor activation in the vestibular primary-afferent neurons of the rat. Neurosci Lett 468: 238–242.
80. Blandizzi C, De Paolis B, Colucci R, Lazzeri G, Baschiera F, et al. (2001) Characterization of a novel mechanism accounting for the adverse cholinergic effects of the anticancer drug irinotecan. Br J Pharmacol 132: 73–84.
81. Yasuda K, Miyake K, Horikawa Y, Hara K, Osawa H, et al. (2008) Variants in *KCNQ1* are associated with susceptibility to type 2 diabetes mellitus. Nat Genet 40: 1092–1097.
82. Ichihara S, Toyooka S, Fujiwara Y, Hotta K, Shigematsu H, et al. (2007) The impact of epidermal growth factor receptor gene status on gefitinib-treated Japanese patients with non-small-cell lung cancer. Int J Cancer 120: 1239–1247.
83. Seo S, Takayama K, Uno K, Ohi K, Hashimoto R, et al. (2013) Functional Analysis of Deep Intronic SNP rs13438494 in Intron 24 of *PCLO* Gene. PLoS One 8: e76960.
84. Sauna ZE, Kimchi-Sarfaty C, Ambudkar SV, Gottesman MM (2007) Silent polymorphisms speak: how they affect pharmacogenomics and the treatment of cancer. Cancer Res 67: 9609–9612.
85. Capon F, Allen MH, Ameen M, Burden AD, Tillman D, et al. (2004) A synonymous SNP of the corneodesmosin gene leads to increased mRNA stability and demonstrates association with psoriasis across diverse ethnic groups. Hum Mol Genet 13: 2361–2368.
86. Nackley AG, Shabalina SA, Tchivileva IE, Satterfield K, Korchynskyi O, et al. (2006) Human catechol-*O*-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. Science 314: 1930–1933.
87. Nielsen KB, Sorensen S, Cartegni L, Corydon TJ, Doktor TK, et al. (2007) Seemingly neutral polymorphic variants may confer immunity to splicing-inactivating mutations: a synonymous SNP in exon 5 of *MCAD* protects from deleterious mutations in a flanking exonic splicing enhancer. Am J Hum Genet 80: 416–432.
88. Spasovski V, Tosic N, Nikcevic G, Stojiljkovic M, Zukic B, et al. (2013) The influence of novel transcriptional regulatory element in intron 14 on the expression of Janus kinase 2 gene in myeloproliferative neoplasms. J Appl Genet 54: 21–26.
89. Xue G, Aida Y, Onodera T, Sakudo A (2012) The 5′ flanking region and intron1 of the bovine prion protein gene (*PRNP*) are responsible for negative feedback regulation of the prion protein. PLoS One 7: e32870.
90. Yoshida T, Ono H, Kuchiba A, Saeki N, Sakamoto H (2010) Genome-wide germline analyses on cancer susceptibility and GeMDBJ database: Gastric cancer as an example. Cancer Sci 101: 1582–1589.

Japanese Journal of Clinical Oncology

Japanese Journal of Clinical Oncology, 2015, 45(3) 307–310
doi: 10.1093/jjco/hyu218
Advance Access Publication Date: 12 January 2015
Short Communication



Short Communication

Relationship between smoking and multiple colorectal cancers in patients with Japanese Lynch syndrome: a cross-sectional study conducted by the Japanese Society for Cancer of the Colon and Rectum

Kohji Tanakaya[1], Yoichi Furukawa[2], Yusuke Nakamura[3], Keiji Hirata[4], Naohiro Tomita[5], Kazuo Tamura[6], Kokichi Sugano[7], Chikashi Ishioka[8], Teruhiko Yoshida[9], Hideyuki Ishida[10,*], Toshiaki Watanabe[11], and Kenichi Sugihara[12] for HNPCC registry and genetic testing project of the Japanese Society for Cancer of the Colon and Rectum

[1]Department of Surgery, Iwakuni Clinical Center, Iwakuni, [2]Division of Clinical Genome Research, Advanced Clinical Research Center, Institute of Medical Science, The University of Tokyo, Tokyo, [3]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, [4]Department of Surgery 1, University of Occupational and Environmental Health, Kitakyushu, [5]Department of Surgery, Hyogo Medical University, Nishinomiya, [6]Major in Science, Graduate School of Science and Engineering Research, Kinki University, Higashio-saka, [7]Oncogene Research Unit/Cancer Prevention Unit, Tochigi Cancer Center, Utsunomiya, [8]Department of Clinical Oncology, Institute of Development, Aging and Cancer, Tohoku University, Sendai, [9]Genetics Division, National Cancer Center Research Institute, Tokyo, [10]Department of Digestive Tract and General Surgery, Saitama Medical Center, Saitama Medical University, Kawagoe, [11]Department of Surgical Oncology, The University of Tokyo, Tokyo, and [12]Department of Surgical Oncology, Graduate School, Tokyo Medical and Dental University, Tokyo, Japan

*For reprints and all correspondence: Hideyuki Ishida, Department of Digestive Tract and General Surgery, Saitama Medical Center, Saitama Medical University, Kamoda, Kawagoe City, Saitama 350-8550, Japan. E-mail: 05hishi@saitama-med.ac.jp

Received 9 September 2014; Accepted 13 December 2014

Abstract

The positive correlation between smoking and cancer risk is well estimated in sporadic colorectal cancer, whereas little is known with regard to Lynch syndrome-associated colorectal cancer. A total of 118 familial colorectal cancer patients from the Hereditary Nonpolyposis Colorectal Cancer Registry and Genetic Testing Project of the Japanese Society for Cancer of the Colon and Rectum, were assessed to determine whether smoking alters the incidence of multiple colorectal cancers. In male patients with Lynch syndrome ($n = 29$), the incidence of multiple colorectal cancers in patients who had ever smoked (smoking duration: median of 19 years) was higher than that in those who never smoked (58.8% vs. 10.0%, $P = 0.02$). The cumulative risk for metachronous colorectal cancer was significantly higher in male Lynch syndrome patients who had previously smoked than in those who had never smoked ($P = 0.03$). Our data suggest that long-term cigarette smoking might be a strong risk factor for the development of multiple colorectal cancers in male Lynch syndrome patients.

Key words: Lynch syndrome, smoking, colorectal cancer, multiple cancer, male

© The Author 2015. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Introduction

Lynch syndrome is among the most common hereditary cancer syndromes, and is caused by the germline mutation of one of the mismatch repair genes, including *MLH1*, *MSH2*, *MSH6* or *PMS2* (1,2). This syndrome accounts for ~1–4% of all colorectal cancer (CRC) patients, and increases the risk of developing CRC, as well as of extra-colonic cancers, including endometrial, ovarian and stomach cancers (1,2). The tumor spectrum of Lynch syndrome has changed over time. The spectrum observed in the first Lynch syndrome family reported by Warthin (3) included mainly gastric cancers and endometrial cancer. However, follow-up reports of this family showed that in current generations, CRC is now the most common tumor. This change suggests that the cancer spectrum of Lynch syndrome is influenced by environmental factors (2). In sporadic CRC, environmental factors such as dietary and lifestyle habits, smoking, alcohol consumption and obesity interact with a host's genetic factors, and may modulate the risk of CRC (4,5). However, there have been only limited reports on the association between smoking and CRC development in patients with Lynch syndrome (6–8). Our objective was to evaluate whether cigarette smoking would alter the age of first CRC onset, incidence of multiple CRCs or occurrence of extra-colonic cancers in individuals with Lynch syndrome.

Patients and methods

This nationwide Japanese study was conducted by the Hereditary Nonpolyposis Colorectal Cancer Registry and Genetic Testing Project of the Japanese Society for Cancer of the Colon and Rectum (9) between September 2002 and July 2010. To determine the patient eligibility, a modified version of the Amsterdam II criteria was used, in which gastric cancer was included as an extra-colonic cancer. All participants provided written informed consent. Clinical information, such as the personal and family cancer history, smoking habits, alcohol use, body height and body weight, were collected either from medical records or directly from patients who were provided genetic counseling. The Brinkman index was calculated as the product of the number of cigarettes smoked per day times the number of years of smoking (9). The body mass index [BMI, (weight, kg)/(height, m^2)] was calculated from the height and weight. Participants with a BMI ≥ 25.0 kg/m^2 were classified as overweight.

Mutation analyses were performed by direct sequencing of the entire coding region in the *MLH1*, *MSH2* and *MSH6* genes. Large deletions/duplications in the *MLH1* and *MSH2* genes were also investigated by multiplex ligation-dependent probe amplification. Participants were stratified into four groups based on gender and the genetic testing results. For each group, we assessed whether smoking altered the age of first CRC onset, incidence of multiple CRCs or the occurrence of extra-colonic cancers. Metachronous CRC was defined as primary CRC diagnosed >12 months after the first diagnosis of primary colon cancer.

The data were calculated as totals, medians (range) or percentages. Fisher's exact test, Mann–Whitney *U*-test and the log-rank test were used for the statistical analyses, as appropriate. The statistical analyses were carried out using JMP 10.0.2 (SAS Institute, Cary, NC); a value of $P < 0.05$ was considered to be statistically significant.

Results

A total of 125 CRC patients were eligible to enter the study, and clinical information was available for review in 118 of these cases. Of the 118 CRC patients, 29 (24.6%) were male Lynch syndrome patients, 34 (28.8%) were female Lynch syndrome patients, 27 (22.9%) were male familial CRC cases [negative for mismatch repair (MMR) genetic testing] and 28 (23.7%) were female familial CRC cases. The male to female ratio of these 118 patients was about three times higher in those who had ever smoked than in those who had never smoked. When the male and female LS patients were analyzed together, there was no significant difference in the incidence of multiple CRCs between patients who had a history of smoking ($n = 25$) and those who had never smoked (44% vs. 29%, $P = 0.22$). When the male and female LS were analyzed separately, the incidence of multiple CRCs in patients who had a history of smoking tended to be lower (17% vs. 57%, $P = 0.15$), and the rate of smoking was significantly lower (18% vs. 66%, $P < 0.01$) in female LS patients than in male LS patients. In addition, the median Brinkman Index of female LS patients who had ever smoked was approximately one-third of that seen in the male LS patients who had ever smoked (170 vs. 550, $P < 0.01$). Thus, we proceeded with a further analysis, focusing on the relationship between multiple CRCs and smoking in the male LS patients.

In the male Lynch syndrome patients, no significant differences were observed in the mismatch repair mutation (*MLH1* in 11 patients and *MSH2* in 18 patients) ($P = 0.33$), age at last follow-up ($P > 0.99$), follow-up period from the first CRC ($P > 0.99$), alcohol consumption ($P = 0.47$) and BMI ($P > 0.99$) between those who had ever smoked ($n = 18$) and those who had never smoked ($n = 9$). The age at which smokers started smoking was 20.0 year (18–30), the years of smoking was 19 (1–44) and the Brinkman index was 550 (75–2300). Among the 19 who had smoked, 57.9% were current smokers and 42.1% were former smokers. The incidence of multiple CRCs in male Lynch syndrome patients who had smoked was 58.8%, compared with 10% in those who had never smoked. Smoking significantly increased the risk ($P = 0.02$) of male Lynch syndrome patients developing multiple CRCs (Table 1). The Brinkman index was applied with a

Table 1. The background characteristics and the cancer risk of the male patients with Lynch syndrome according to the smoking status

	Ever smoked ($n = 19$)	Never smoked ($n = 10$)	*P* value
Mismatch repair gene mutation (*MLH1/MSH2/MSH6*)	6/13/0	5/5/0	0.43
Age at the last follow-up (years)	56 (27–78)	47 (28–74)	>0.99
Follow-up period from first CRC (years)	13.3 (0.2–28)	6.5 (0.3–22.4)	>0.99
Consumption of alcohol ≥3 days/week	10/18 (55.6%)	4/10 (40.0%)	0.47
BMI (kg/m^2) ≧ 25.0	5/17 (29.4%)	2/9 (22.2%)	>0.99
Age at first CRC (years)	42 (24–70)	34.5 (25–66)	0.69
Multiple CRC	10/17 (58.8%)	1/10 (10.0%)	0.02
Metac/Sync/Metac + Sync	8/1/1	0/1/0	
Extra-colonic cancer	5/18 (27.7%)	4/10 (40.0%)	>0.99

CRC, colorectal cancer; BMI, body mass index; Sync, synchronous; Metac, metachronous.

cut-off of ≥400, which resulted in an area under the receiver operating characteristic *curve* of 0.84, and identified multiple CRCs with 81.8% sensitivity and 76.5% specificity. A Kaplan–Meier plot of the metachronous CRC-free rates showed that the cumulative risk for metachronous CRC was significantly higher in male Lynch syndrome patients who had previously smoked than in those who had never smoked ($P = 0.03$, log rank) (Fig. 1).

We observed no significant positive trend in multiple CRC risk with an increasing duration of smoking or an increasing amount smoked. Smoking was not associated with either the age at first CRC onset ($P = 0.69$) or the occurrence of extra-colonic cancers ($P > 0.99$) in male Lynch syndrome patients. The observed extra-colonic cancers in male Lynch syndrome patients included three gastric cancers, one renal pelvis cancer, one urinary bladder cancer and one prostate cancer in five patients with a history of smoking, whereas two gastric cancers, one duodenal cancer, one urinary bladder cancer and one thigh liposarcoma were observed in four patients who had no history of smoking. In the other groups, smoking was not associated with the age at the first CRC onset, the incidence of multiple CRCs, or the occurrence of extra-colonic cancers.

Discussion

Colorectal cancer is one of the most common cancers in Western countries, and its incidence has increased recently in Eastern Asia, especially in Japan, as it has become westernized over the past few decades (10). In fact, the high incidence in Japanese immigrants to Hawaii suggests that the changes in environmental factors, including the westernization of dietary habits and lifestyle, may contribute to this increase (4). Moreover, according to the Asia Pacific Cohort Studies Collaboration, changes in lifestyle, such as smoking, the BMI and a lack of physical activity, increase the risk of CRC (5). Many studies have reported a 20–60% increase in the risk of CRC associated with smoking (11). Cigarette smoke contains numerous carcinogens and toxicants. A strong relationship between tobacco smoking, carcinogen–DNA adduct formation, smoke exposure and cancer risk has been demonstrated.

Individuals with Lynch syndrome have a very high risk of developing CRC (25–70%) and endometrial cancer (30–70%), and an increased risk of developing other tumors (12). The main clinical features are an early age of onset and the occurrence of multiple tumors. For Lynch syndrome patients who undergo segmental resection of their initial cancer, the cumulative risk of metachronous CRC is 16% at 10 years, 41% at 20 years and up to 62% after 30 years (13). Thus, in patients with Lynch syndrome, managing the risk of developing CRC is very important.

There have been many studies performed to determine which environmental and lifestyle factors influence the development of CRC in patients with Lynch syndrome. Two studies demonstrated that a higher BMI was associated with an increased risk of colorectal neoplasia (12). Alcohol was not found to be associated with the risk, while fruit and fiber intake were found to be possibly related to decreased risk (12). Concerning smoking, two studies (one retrospective cohort (6) and another case–control (8) study) showed a significant association between smoking and an increased risk of developing CRC in patients with Lynch syndrome. The retrospective study (6) showed that, for former smokers, the CRC risk decreased as the number of years since quitting increased ($P < 0.01$).

Our data suggest that long-term smoking (median of 19 years) significantly increased the risk of developing multiple CRCs in male patients with Lynch syndrome. We believe that intensive colorectal surveillance and smoking cessation are very important in patients with Lynch syndrome, especially in males. One limitation is the small sample size and lack of standardization for the surveillance program. Besides smoking, we tried to identify any other risk factors for developing CRC. However, we could not identify any significant association between alcohol consumption, BMI, MMR gene mutation and multiple CRC risk. The other limitation was a lack of sufficient data regarding the effect of smoking on the development of multiple CRCs in female LS syndrome patients, because the smoking habits of females are generally different from those of males in Japan, which also seems to be applicable to female LS patients. The relationship between smoking and multiple CRCs in female LS patients is considered to be important, but further investigations with larger series are needed.

The present results obtained from male LS syndrome patients also need further confirmation, preferably in prospective studies. However, smoking is also a major cause of many other deadly health problems, such as heart disease, aneurysms, bronchitis, emphysema and stroke. Although inconclusive data are available about reducing the risk of developing multiple CRCs at this time, the recommendation of smoking cessation is a reasonable option for all patients, including males with Lynch syndrome, as it can help prevent the development of various conditions.

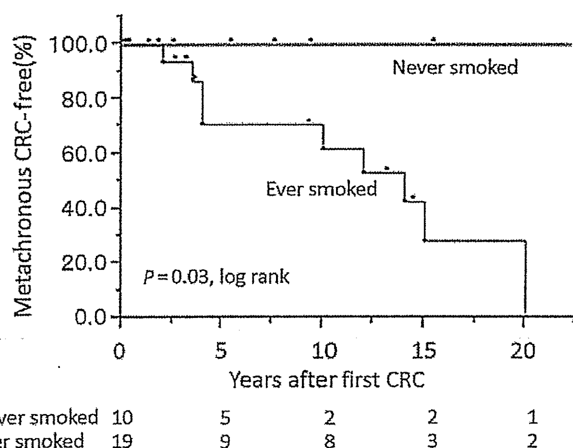


Figure 1. The Kaplan–Meier plots of the metachronous colorectal cancer-free rates by smoking status in male patients with Lynch syndrome. CRC, colorectal cancer.

Conclusions

Our data suggest that cigarette smoking might be a strong risk factor for the development of multiple CRCs after long-term exposure in male patients with Lynch syndrome.

Funding

The present study was supported in part by a grant-in-aid for Cancer Research from the Ministry of Health, Labor and Welfare, and by the Japanese Society for Cancer of the Colon and Rectum.

Conflict of interest statement

None declared.

References

1. Lynch HT, de la Chapelle A. Hereditary colorectal cancer. *N Engl J Med* 2003;348:919–32.
2. Van Duijnhoven FJB, Botma A, Winkels R, Nagengast FM, Vasen HFA, Kampman E. Do lifestyle factors influence colorectal cancer risk in Lynch syndrome? *Familial Cancer* 2013;12:285–93.
3. Warthin AS. Hereditary with reference to carcinoma. *Arch Intern Med* 1913;12:546–55.
4. Otani T, Iwasaki M, Yamamoto S, et al. Alcohol consumption, smoking, and subsequent risk of colorectal cancer in middle-aged and elderly Japanese men and women: Japan Public Health Center-based prospective study. *Cancer Epidemiol Biomark Prev* 2003;12:1492–500.
5. Huxley R. Asia Pacific Cohort Studies Collaboration. The role of lifestyle risk factors on mortality from colorectal cancer in populations of the Asia-Pacific region. *Asian Pac J Cancer Prev* 2007;8:191–8.
6. Pande M, Lynch PM, Hopper JL, et al. Smoking and colorectal cancer in Lynch syndrome: results from the Colon Cancer Family Registry and the University of Texas M.D. Anderson Cancer Center. *Clin Cancer Res* 2010;16:1331–9.
7. Brand RM, Jones DD, Lynch HT, et al. Risk of colon cancer in hereditary non-polyposis colorectal cancer patients as predicted by fuzzy modeling: Influence of smoking. *World J Gastroenterol* 2006;12:4485–91.
8. Diergaarde B, Braam H, Vasen HF, et al. Environmental factors and colorectal tumor risk in individuals with hereditary nonpolyposis colorectal cancer. *Clin Gastroenterol Hepatol* 2007;5:736–42.
9. Brinkman GL, Coates EO Jr. The effect of bronchitis, smoking and occupation on ventilation. *Ann Respir Dis* 1963;87:684–93.
10. Katanoda K, Matsuda T, Matsuda A, et al. An updated report of the trends in cancer incidence and mortality in Japan. *Jpn J Clin Oncol* 2013;43:492–507.
11. Hannan LM, Jacobs EJ, Thun MJ. The association between cigarette smoking and risk of colorectal cancer in a large prospective cohort from the United States. *Cancer Epidemiol Biomark Prev* 2009;18:3362–7.
12. Vasen HFA, Ignacio B, Aktan-Collan K, et al. Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. *Gut* 2013;62:812–23.
13. Parry S, Win AK, Parry B, et al. Metachronous colorectal cancer risk for mismatch repair gene mutation carriers: the advantage of more extensive colon surgery. *Gut* 2011;60:950–7.

Appendix

In addition to the authors listed in the author field, following authors have contributed equally to the study.

Tatsuro Yamaguchi: Department of Surgery, Tokyo Metropolitan Cancer and Infectious diseases Center Komagome Hospital, Tokyo, Japan.

Hideki Ishikawa: Department of Molecular-Targeting Cancer Prevention and Epidemiology for Community Health and Medicine, Kyoto Prefectural University of Medicine, Kyoto, Japan.

Nagahide Matsubara: Department of Surgery, Hyogo Medical University, Nishinomiya, Japan.

Masami Arai: Clinical Genetic Oncology, The Cancer Institute Hospital of Japanese Foundation for Cancer Research, Tokyo, Japan;

Yoshihiro Moriya: Department of Surgery, National Cancer Center Research Institute, Tokyo, Japan.