J Hum Genet (2003) 48:654–658
DOI 10.1007/s10038-003-0096-1

ORIGINAL ARTICLE

**Kimio Yoshimura · Tomoyuki Hanaoka
Shumpei Ohnami · Sumiko Ohnami · Takashi Kohno
Ying Liu · Teruhiko Yoshida · Hiromi Sakamoto
Shoichiro Tsugane**

Allele frequencies of single nucleotide polymorphisms (SNPs) in 40 candidate genes for gene-environment studies on cancer: data from population-based Japanese random samples

Received: 11 July 2003 / Accepted: 1 October 2003 / Published online: 21 November 2003

Abstract Knowledge of genetic polymorphisms in gene-environment studies may contribute to more accurate identification of avoidable risks and to developing tailor-made preventative measures. The aim of this study was to describe the allele frequencies of single nucleotide polymorphisms (SNPs) of select genes, which may be included in future gene-environment studies on cancer in Japan. SNP typing was performed on middle-aged Japanese men randomly selected from the general population in five areas of Japan. We genotyped and calculated allele frequencies of 153 SNPs located on 40 genes: CYP1A1, CYP1B1, CYP2C9, CYP2C19, CYP2E1, CYP17A1, CYP19A1, AHR, ESR1, ESR2, ERRRG, PGR, EPHX1, EPHX2, HSD17B2, HSD17B3, GSTM2, GSTM3, GSTT2, GSTP1, NAT1, NAT2, COMT, ADH1A, ADH1B, ADH1C, ALDH2, NOS2A, NOS3, IL1A, IL1B, OGG1, NUDT1 [MTH1], DRD2, DRD3, DRD4, SLC6A4, NR3C1 [GCCR], MTHFR, and NQO1. In the present study, the Japanese allele frequencies were verified by using nationwide population samples.

Keywords Gene frequency · Single nucleotide polymorphism · Japanese population · Gene-environment study · Cytochrome P-450 enzyme system · DNA repair enzymes

K. Yoshimura (✉) · Y. Liu
Cancer Information and Epidemiology Division,
National Cancer Center Research Institute,
5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan
E-mail: kyoshimu@ncc.go.jp
Tel.: +81-3-35422511 ext. 4254
Fax: +81-3-35460630

T. Hanaoka · S. Tsugane
Epidemiology and Biostatistics Division,
National Cancer Center Research Institute East, Chiba, Japan

S. Ohnami · S. Ohnami · T. Yoshida · H. Sakamoto
Genetics Division, National Cancer Center Research Institute,
Tokyo, Japan

T. Kohno
Biology Division, National Cancer Center Research Institute,
Tokyo, Japan

Introduction

Both environmental and host susceptibility factors are relevant to the etiology of nonhereditary cancers. Although it has been difficult to estimate the host susceptibility factors accurately and comprehensively in cancer epidemiology, analysis of genetic polymorphism offers a new opportunity to address this difficulty. Information on genetic differences of genes that may modify dose dependency or effects of carcinogenic exposure will contribute to more accurate identification of which risks can be intervened, and to establish tailor-made prevention measures.

The objective of this study was to describe the allele frequencies of single nucleotide polymorphisms (SNPs) of selected genes, which may be important in future gene-environment studies in Japan. Genes were selected based on their possible involvement in gene-environment/life-style interactions and constitute the genes of xenobiotic metabolism enzymes, DNA repair enzymes, and other stress-related proteins.

Although many studies have reported allele frequencies of various SNPs in Japan, little has been published using a strict random sampling method in the subject selection. Here we report allele frequencies of a large number of SNPs, of which typing was performed on samples obtained from middle-aged Japanese men randomly selected from areas served by five public health centers in different prefectures.

Subjects and methods

Subjects

Selection of five study areas and subjects was described previously (Tsugane et al. 1992a; Tsugane et al. 1992b). Briefly, each study

Table 1 Single nucleotide polymorphisms (SNP) allele frequencies

Gene	SNP (rs no.)	Reference/ variant allele	Amino acid change	Amino acid number	Sample size						Major/ minor allele	Minor allele frequency	(95% CI)
					Akita	Iwate	Nagano	Okinawa	Tokyo	Total			
CYP1A1	rs1048943	gacc(A/G)ttgc	Ile/Val	462	75	77	104	43	35	334	A/G	0.214	(0.184–0.245)
CYP1B1	rs10012	gctc(C/G)ggtc	Arg/Gly	48	77	75	104	45	35	336	C/G	0.098	(0.075–0.121)
	rs1056827	gccg(G/T)cctt	Ala/Ser	119	77	77	103	38	33	328	G/T	0.099	(0.076–0.122)
	rs1056836	ccca(C/G)tgaa	Leu/Val	432	77	76	104	44	36	337	C/G	0.160	(0.132-0.188)
	rs10916	atga(T/G)ttat			76	75	104	43	34	332	T/G	0.140	(0.113–0.167)
CYP2C9	rs1505	gatt(G/C)agag			69	69	83	32	26	279	G/C	0.486	(0.442–0.529)
CYP2C19	rs1322179	tagg(G/A)ctgg			77	77	104	45	35	338	G/A	0.303	(0.269–0.337)
CYP2E1	rs3813867	agag(G/C)tgca			77	76	104	44	35	336	G/C	0.228	(0.195–0.261)
	rs2031920	agta(C/T)aaaa			75	68	96	32	27	298	C/T	0.227	(0.192–0.261)
	rs2070673	cagg(T/A)cggt			77	76	103	37	33	326	T/A	0.445	(0.405–0.484)
CYP17A1	rs6162	gaca(T/C)ggcc	His/His	46	75	69	95	31	27	297	C/T	0.449	(0.410–0.489)
	rs6163	attc(T/G)gttc	Ser/Ser	65	77	76	102	37	34	326	G/T	0.449	(0.411–0.487)
	rs743572	ccac(T/C)gctg			77	77	104	45	35	338	T/C	0.451	(0.414–0.489)
CYP19A1	rs700518	cata(T/C)accc	Val/Val	80	77	77	104	45	36	339	T/C	0.326	(0.291–0.361)
	rs700519	ctgc(G/A)tctt	Arg/Cys	264	77	77	104	45	35	338	G/A	0.330	(0.295–0.365)
	rs4646	tgac(A/C)aata			74	65	93	22	22	276	C/A	0.283	(0.244–0.321)
AHR	rs2066853	gtgt(C/T)tgat	Arg/Lys	554	77	73	97	33	31	311	C/T	0.450	(0.410–0.490)
	rs713150	tata(C/G)ttct			70	66	94	28	26	284	G/C	0.280	(0.243–0.317)
	rs2074113	acat(C/A)caat			73	53	86	19	17	248	C/A	0.448	(0.402–0.494)
	rs2237297	tgct(C/T)ttgt			77	77	104	45	34	337	C/T	0.438	(0.400–0.476)
	rs2237299	gatt(A/G)gtga			77	77	104	45	34	337	A/G	0.356	(0.320–0.392)
	rs2282886	cttt(T/C)ctat			77	77	104	45	36	339	T/C	0.357	(0.321–0.392)
ESR1	rs1913474	gttc(T/C)aaga			77	76	103	37	32	325	C/T	0.465	(0.426–0.503)
	rs932479	tatc(C/T)ctca			76	77	104	45	34	336	T/C	0.469	(0.432–0.506)
	rs2011885	ttga(G/A)catt			75	65	92	20	20	272	A/G	0.493	(0.451–0.534)
	rs974276	tata(A/G)aggt			77	77	104	45	33	336	A/G	0.302	(0.267–0.337)
	rs1062577	attc(T/A)tttt			77	75	103	37	30	322	T/A	0.318	(0.281–0.356)
ESR2	rs1256054	gttg(G/C)agtt	Leu/Leu	392	75	70	95	33	28	301	G/C	0.046	(0.028–0.064)
	rs1256049	gccg(C/T)actt	Val/Val	328	75	70	95	33	30	303	C/T	0.256	(0.220–0.291)
	rs1256027	tact(G/A)tcca			77	77	103	38	32	327	G/A	0.255	(0.221–0.290)
	rs944459	agta(C/T)gcaa			77	77	103	38	33	328	C/T	0.165	(0.136–0.194)
	rs2274705	aaat(A/C)tatg			77	75	102	36	33	323	A/C	0.086	(0.064–0.109)
	rs1256030	tggg(A/G)gcta			77	77	103	38	33	328	G/A	0.416	(0.378–0.454)
ESRRG	rs1498283	gggt(G/A)agaa			76	77	104	45	34	336	A/G	0.424	(0.385–0.463)
	rs1339343	taac(T/C)cagt			75	68	99	31	27	300	T/C	0.076	(0.054–0.098)
PGR	rs484389	gtgg(C/T)atgt			77	77	103	45	34	336	T/C	0.188	(0.159–0.216)
EPHX1	rs1051741	ccaa(T/C)gtca	Asn/Asn	357	77	77	104	45	36	339	C/T	0.177	(0.149–0.205)
	rs1051740	caga(C/T)accc	His/Tyr	113	76	74	98	43	35	326	T/C	0.400	(0.361–0.440)
	rs6965	cacc(A/G)ctgc			76	77	103	45	36	337	A/G	0.285	(0.250–0.319)
	rs2292566	ctaa(G/A)attg	Lys/Lys	119	77	77	104	45	36	339	G/A	0.313	(0.279–0.347)
	rs2234922	ggcc(G/A)tacc	His/Arg	139	77	75	102	34	34	322	A/G	0.183	(0.155–0.212)
	rs2292568	accc(C/T)gtca	Pro/Pro	284	77	75	102	34	34	322	C/T	0.132	(0.106–0.158)
EPHX2	rs751141	gacc(C/T)ggta	Arg/Gln	287	77	75	101	34	33	320	C/T	0.195	(0.163–0.227)
	rs1042032	agga(T/C)ggca			76	77	104	45	35	337	T/C	0.451	(0.413–0.489)
	rs891401	aata(C/G)catg			77	77	102	45	35	336	C/G	0.268	(0.233–0.303)
	rs2291635	tgca(C/T)aaca			77	77	104	45	35	338	C/T	0.169	(0.140–0.197)
	rs1126452	cggt(T/G)ggcc	Pro/Pro	531	77	77	104	45	34	337	T/G	0.461	(0.424–0.499)
	rs747276	aggt(C/G)atgg			77	74	100	34	30	315	C/G	0.273	(0.237–0.309)
HSD17B2	rs2042429	aaat(T/C)aatg			77	74	100	34	33	318	C/T	0.239	(0.206–0.272)
	rs1017243	ttct(A/G)ggac			77	75	102	34	34	322	G/A	0.200	(0.168–0.232)
	rs1424151	atcc(A/G)taga			77	74	99	34	32	316	A/G	0.117	(0.091–0.142)
	rs996752	gaca(A/G)tgct			75	69	94	32	28	298	A/G	0.292	(0.256–0.327)
HSD17B3	rs2066480	atgt(G/A)tttt	Val/Ile	31	77	77	104	45	36	339	G/A	0.045	(0.029–0.061)
	rs375944	gtat(T/G)agtc			74	69	94	34	28	299	T/G	0.286	(0.250–0.322)
	rs280654	gtct(T/C)ttgc			77	77	104	45	36	339	T/C	0.267	(0.234–0.300)
	rs912462	aagc(C/T)gtct			71	65	87	23	21	267	T/C	0.154	(0.122–0.185)
	rs2066479	cagc(G/A)gtgc	Gly/Ser	289	77	77	104	45	35	338	G/A	0.262	(0.228–0.295)
	rs867807	agta(A/G)actt			77	77	104	45	34	337	A/G	0.017	(0.007–0.027)
GSTM2	rs655315	gaca(G/A)aaga			75	63	89	19	19	265	G/A	0.258	(0.220–0.297)
	rs428434	atga(G/C)aaat			71	59	89	18	20	257	G/C	0.230	(0.194–0.265)
GSTM3	rs1332018	atgt(C/A)gggt			75	68	98	31	26	298	A/C	0.191	(0.161–0.221)
GSTT2	rs1622002	ccat(G/A)gacc	Met/Ile	139	77	75	99	34	32	317	G/A	0.159	(0.131–0.188)
	rs2719	tgct(G/T)ctac			77	77	104	45	35	338	G/T	0.232	(0.200–0.264)
	rs2267047	ataa(C/T)gtta			77	77	104	45	36	339	C/T	0.389	(0.354–0.425)
	rs140186	ttag(G/A)ggat			75	70	95	32	27	299	G/A	0.378	(0.339–0.417)

Table 1 (Continued)

Gene	SNP (rs no.)	Reference/ variant allele	Amino acid change	Amino acid number	Sample size						Major/ minor allele	Minor allele frequency	(95% CI)
					Akita	Iwate	Nagano	Okinawa	Tokyo	Total			
GSTP1	rs947894	atac(A/G)tctc	Ile/Val	105	77	77	104	45	35	338	A/G	0.151	(0.124–0.178)
	rs762803	tgaa(C/A)ctcc			77	76	102	44	36	335	C/A	0.146	(0.119–0.174)
NAT1	rs15561	atgt(C/A)tttt			72	51	77	18	11	229	A/C	0.439	(0.392–0.485)
NAT2	rs1801280	acca(T/C)tgac	Ile/Thr	114	72	54	78	17	9	230	T/C	0.010	(0.001–0.020)
	rs1799929	gtac(C/T)tgga	Leu/Leu	161	73	54	78	17	10	232	C/T	0.010	(0.001–0.020)
	rs1799930	cctc(G/A)aaca	Arg/Gln	197	70	54	78	17	6	225	G/A	0.178	(0.143–0.213)
	rs1495744	atga(A/G)ttaa			74	70	94	30	26	294	A/G	0.011	(0.003–0.020)
COMT	rs4633	acca(C/T)gtgc	His/His	62	69	52	81	16	9	227	C/T	0.286	(0.245–0.328)
	rs4680	tggc(G/A)tgaa	Val/Met	158	77	77	104	45	35	338	G/A	0.322	(0.287–0.358)
	rs6267	gaac(G/T)caca	Ala/Ser	72	76	75	102	44	34	331	G/T	0.103	(0.079–0.126)
	rs2097603	ggaa(G/A)cttt			72	55	77	7	8	219	A/G	0.295	(0.252–0.337)
	rs2020917	tggc(C/T)agtg			77	77	104	45	35	338	C/T	0.257	(0.225–0.290)
	rs2239393	cccc(A/G)tttc			77	77	104	45	35	338	A/G	0.250	(0.218–0.282)
ADH1A	rs931635	acaa(G/A)ctaa			70	54	76	7	6	213	G/A	0.093	(0.066–0.121)
	rs1229967	tagc(C/G)tcca			73	54	80	21	10	238	C/G	0.098	(0.072–0.125)
	rs1229970	aaat(C/A)tatt			75	56	88	19	11	249	C/A	0.088	(0.064–0.113)
	rs975833	ataa(G/C)cata			75	56	83	19	11	244	C/G	0.170	(0.139–0.201)
	rs1618572	agct(G/C)tttc			72	52	77	7	8	216	G/C	0.090	(0.063–0.117)
	rs2276332	gcct(A/C)aatg			74	56	81	19	11	241	A/C	0.062	(0.040–0.083)
ADH1B	rs17033	atct(T/C)gtag			72	53	71	8	8	212	T/C	0.066	(0.042–0.089)
	rs1159918	agac(A/C)ccca			71	54	75	8	8	216	C/A	0.160	(0.127–0.193)
	rs1042026	ctct(T/C)ggac			66	52	73	7	8	206	C/T	0.167	(0.134–0.201)
ADH1C	rs1789924	gtta(T/C)gaag			72	55	77	7	8	219	C/T	0.057	(0.036–0.078)
	rs1693430	aaat(T/C)ggtg			71	55	75	7	8	216	C/T	0.055	(0.034–0.076)
	rs2009181	tttg(A/G)tttt			71	54	77	7	8	217	G/A	0.124	(0.094–0.155)
	rs2298755	cttt(G/C)acaa			75	54	80	19	11	239	C/G	0.064	(0.042–0.086)
	rs3216150	aaaa(A/-)tcac			74	54	81	19	9	237	*/A	0.065	(0.043–0.087)
ALDH2	rs671	cact(G/A)aagt	Glu/Lys	504	77	69	98	45	35	324	G/A	0.171	(0.142–0.201)
	rs2238151	cctg(A/G)gcca			77	77	103	45	36	338	G/A	0.074	(0.054–0.093)
	rs2238152	caaa(C/A)agat			77	77	104	45	34	337	C/A	0.151	(0.125–0.178)
	rs441	tgag(A/G)ccga			77	76	104	45	36	338	A/G	0.151	(0.124–0.177)
NOS2A	rs1060826	tctc(T/C)gtgg	Thr/Thr	919	64	55	76	7	3	205	C/T	0.185	(0.147–0.224)
	rs1060822	ggat(A/G)cctt	Gly/Gly	786	71	55	77	7	8	218	G/A	0.209	(0.170–0.247)
	rs2072324	ttat(C/A)ttct			76	76	102	45	36	335	C/A	0.303	(0.268–0.338)
	rs2297518	gagc(G/A)attt	Ser/Leu	608	66	55	70	14	8	213	G/A	0.049	(0.029–0.069)
	rs2297520	ccgg(T/C)ggct			77	77	104	45	36	339	C/T	0.478	(0.440–0.516)
NOS3	rs1800783	caga(T/A)ccca			76	72	95	42	30	306	T/A	0.136	(0.109–0.162)
	rs1549758	agcc(A/G)tcct	Asp/Asp	258	77	77	104	38	34	330	G/A	0.123	(0.097–0.148)
	rs1799983	gggg(A/C)tcat	Asp/Glu	298	77	77	104	43	35	335	C/A	0.074	(0.054–0.095)
	rs1800780	gtcc(T/C)gggt			77	77	103	43	33	328	C/T	0.372	(0.334–0.410)
	rs1800779	tgtg(C/T)catc			77	77	104	43	35	336	T/C	0.129	(0.105–0.154)
IL1A	rs17561	gtca(G/T)cacc	Ala/Ser	114	77	77	104	45	36	339	G/T	0.103	(0.080–0.126)
	rs1800587	aaca(C/T)catt			75	69	97	28	24	293	C/T	0.106	(0.081–0.130)
	rs1800794	gctg(C/T)tttc			77	77	103	45	35	337	C/T	0.101	(0.078–0.123)
	rs2071374	ttgg(A/C)atat			75	74	103	35	33	320	A/C	0.142	(0.116–0.169)
IL1B	rs1143627	aagc(C/T)ataa			77	77	104	43	35	336	T/C	0.418	(0.380–0.456)
	rs16944	cctc(A/G)ggag			76	77	104	45	35	337	G/A	0.415	(0.377–0.454)
	rs1071676	ttaa(G/C)actg			77	76	104	45	34	334	G/C	0.043	(0.028–0.058)
	rs1143637	ttcc(G/A)ctcc			76	75	104	45	36	335	G/A	0.043	(0.028–0.058)
	rs1143629	aaga(C/T)tcca			77	74	104	43	35	335	T/C	0.422	(0.384–0.461)
	rs1143634	tctt(C/T)gaca	Phe/Phe	105	77	77	103	43	35	337	C/T	0.044	(0.029–0.059)
OGG1	rs2075747	ggcc(G/A)cgca			74	66	97	30	26	293	G/A	0.305	(0.269–0.342)
	rs1052133	caat(C/G)ccgc	Ser/Cys	326	77	77	104	45	36	339	C/G	0.445	(0.409–0.482)
	rs2072668	catt(C/G)tgtg			77	77	103	45	36	338	C/G	0.450	(0.413–0.487)
	rs1801129	tact(A/G)cggg			77	77	104	45	33	336	A/G	0.050	(0.033–0.067)
NUDT1	rs4866	tgca(C/T)gtcc	Val/Met	106	75	75	102	44	35	331	C/T	0.103	(0.079–0.126)
	rs1062492	ttgc(G/A)gctg			75	73	97	33	32	310	G/A	0.056	(0.037–0.075)
DRD2	rs1076560	aggg(T/G)gaaa			77	76	104	45	35	335	G/T	0.360	(0.324–0.395)
	rs1124491	gcct(T/C)gctg			77	77	104	43	36	337	C/T	0.353	(0.318–0.388)
	rs6277	ctcc(C/T)gaca	Pro/Pro	319	77	77	104	44	36	339	C/T	0.056	(0.038–0.073)
	rs6275	acca(C/T)ggtc	His/His	313	77	77	103	33	35	335	T/C	0.401	(0.365–0.438)
	rs1076563	accc(T/G)taga			77	77	104	35	35	337	T/G	0.053	(0.036–0.069)
	rs1079596	caac(A/G)ttac			75	70	96	38	30	304	G/A	0.350	(0.312–0.388)
	rs1801028	ccgt(C/G)ccac	Ser/Cys	311	75	71	95	38	30	306	C/G	0.040	(0.025–0.056)
	rs1116313	tacc(T/C)ctct			77	77	104	33	34	330	T/C	0.054	(0.037–0.071)

Table 1 (Continued)

Gene	SNP (rs no.)	Reference/ variant allele	Amino acid change	Amino acid number	Sample size						Major/ minor allele	Minor allele frequency	(95% CI)
					Akita	Iwate	Nagano	Okinawa	Tokyo	Total			
DRD3	rs6280	gagt(A/G)gcca	Ser/Gly	9	77	77	104	33	33	329	A/G	0.269	(0.236–0.302)
	rs1800828	ctca(G/C)agag			76	70	96	38	29	304	G/C	0.255	(0.222–0.288)
DRD4	rs1800955	aggg(T/C)gcgc			66	54	70	14	8	212	T/C	0.370	(0.326–0.414)
	rs936460	cctt(C/T)gagg			74	70	94	30	26	294	T/C	0.121	(0.094–0.147)
	rs752306	ctca(C/T)ctgc			77	77	104	45	35	338	C/T	0.109	(0.085–0.133)
SLC6A4	rs1042173	tata(G/T)aatt			72	54	75	7	8	216	G/T	0.181	(0.144–0.217)
	rs2020939	tacc(A/G)cact			77	77	104	45	36	339	A/G	0.155	(0.127–0.182)
	rs2020936	agcg(A/G)tctc			77	77	104	45	36	339	A/G	0.076	(0.056–0.097)
	rs1872924	gtca(T/C)ctag			74	77	99	44	35	329	T/C	0.074	(0.053–0.095)
	rs25528	tggt(T/G)tcgc			73	77	98	43	34	325	T/G	0.073	(0.053–0.094)
	rs717742	gcgc(A/T)gaca			75	71	95	33	28	302	A/T	0.076	(0.054–0.097)
NR3C1	rs6194	taca(T/C)ctgg	His/His	588	73	77	98	44	35	327	C/T	0.061	(0.042–0.079)
	rs258751	agga(T/C)ggtc	Asp/Asp	678	77	77	104	45	36	339	C/T	0.060	(0.042–0.078)
	rs6196	caaa(C/T)ggaa	Asn/Asn	766	72	55	77	9	8	221	T/C	0.052	(0.031–0.072)
	rs33388	gtga(T/A)taac			75	71	95	33	28	302	A/T	0.159	(0.129–0.189)
	rs33389	actc(A/G)catc			76	77	104	45	34	336	G/A	0.061	(0.042–0.079)
	rs174050	gaag(T/C)cagg			77	73	95	33	31	309	C/T	0.064	(0.045–0.084)
MTHFR	rs2066470	gctc(G/A)gggt	Pro/Pro	39	77	77	104	45	36	339	G/A	0.108	(0.084–0.131)
	rs1801133	ggag(C/T)cgat	Ala/Val	222	77	73	95	33	31	309	C/T	0.401	(0.363–0.439)
	rs1801131	gaag(A/C)aagt	Ala/Glu	429	57	41	75	35	17	225	A/C	0.213	(0.174–0.253)
	rs2066471	gccc(C/T)gaca			77	77	103	45	36	338	C/T	0.105	(0.082–0.128)
	rs2274976	gagc(G/A)gtgg	Arg/Gln	594	77	77	104	45	35	338	G/A	0.105	(0.081–0.129)
NQO1	rs1800566	agaa(C/T)ctca	Pro/Ser	187	77	77	104	43	35	336	C/T	0.378	(0.340–0.416)

area had a population of approximately 100,000 people and was covered by a single health center that supervised the health administration of the several cities, towns, and villages in the area. Men aged 40–49 years were selected by a random sampling method using the publicly available resident registration rolls of the following municipalities: Ninohe, Iwate Prefecture ($n = 175$), Yokote, Akita Prefecture ($n = 170$), Katsushika-kita, Tokyo ($n = 195$), Saku, Nagano Prefecture ($n = 170$), and Ishikawa, Okinawa Prefecture ($n = 170$). These study areas have different characteristic cancer mortality rates (Tsugane et al. 1992a). The selected individuals were sent a letter accompanied by a prepaid reply postcard to explain the purpose of the study and to request their voluntary participation. In order to achieve a sufficient response rate, a follow-up letter, telephone call, and home visit were used to encourage participation.

In addition to the blood and urine samples, the subjects provided information about their life style and health-related condition through a questionnaire-based interview by registered nurses or nutritionists who as local public employees have a professional obligation to strict secrecy under the Local Public Service Law. The overall participation rate was 72% (634 out of 880). The surveys were done in 1989 for Ninohe and Ishikawa, in 1990 for Yokote and Saku, and in 1991 for Katsushika-kita.

DNA samples

A total of 25 ml of blood was drawn by venipuncture and divided into three tubes for extraction of plasma, buffy coat, and serum. The 11-ml heparinized sample was immediately centrifuged for 10 min at 2,500–3,000 rpm to obtain plasma and a buffy coat layer. Genomic DNA was extracted from the buffy coat layer using a commercial kit (Wako, Osaka, Japan). Some samples from Iwate, Okinawa, and Tokyo were used in another study (Sugimura et al. 1998), leaving 537 DNA samples (98 from Akita, 121 from Iwate, 111 from Nagano, 105 from Okinawa, and 102 from Tokyo). A total of 339 extractions yielded 0.5 μg or more DNA and were used for the present study (77 from Akita, 77 from Iwate, 104 from Nagano, 45 from Okinawa, and 36 from Tokyo), but some DNA samples were exhausted before the completion of the analyses, leaving a sample size of 207 (64 from Akita, 54 from Iwate, 67 from Nagano, 14 from Okinawa, and 8 from Tokyo), for which all SNPs were analyzed. Because the substantial variation in the amount of DNA extracted from the buffy coat layer was most likely due to technical reasons and should have nothing to do with the genotype per se, it is reasonable to suppose that random sampling is well preserved in the present study.

Ethical issues

All DNA samples were rendered anonymous by removing links with specific individual information, i.e., any ID, name, or address. The protocol was approved by the ethics review committee of the National Cancer Center (protocol number G12-02).

SNPs analyses

We developed 289 SNP typing assays for 44 genes using a mass spectroscopy-based technique, MassARRAY (Sequenom, CA, USA; Ross et al. 1998).

The following 44 genes were selected from genes encoding xenobiotic metabolic enzymes, DNA repair enzymes, and other stress-related proteins: cytochrome P450 genes (CYP1A1, CYP1A2, CYP1B1, CYP2A6, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP3A4, CYP17A1, CYP19A1), aryl hydrocarbon receptor gene (AHR), estrogen receptor genes (ESR1, ESR2, ERRRG), progesterone receptor gene (PGR), epoxide hydrolase genes (EPHX1, EPHX2), hydroxysteroid (17-beta) dehydrogenase genes (HSD17B2, HSD17B3), glutathione S-transferase genes (GSTM2, GSTM3, GSTT2, GSTP1), N-acetyltransferase genes (NAT1, NAT2), catechol-O-methyltransferase gene (COMT), alcohol dehydrogenase genes (ADH1A, ADH1B, ADH1C), aldehyde dehydrogenase gene (ALDH2), nitric oxide synthase genes (NOS2A, NOS3), interleukin genes (IL1A, IL1B), repair genes for oxidative DNA damage (OGG1, NUDT1 [MTH1]), dopamine receptor genes (DRD2, DRD3, DRD4), serotonin transporter gene (SLC6A4), glucocorticoid receptor gene (NR3C1 [GCCR]), folate metabolizing enzyme gene (MTHFR), and quinone oxidoreductase gene (NQO1). For each gene, between five and seven SNPs were chosen from public databases and published papers.

Results and discussion

Of the 289 SNPs in 44 genes initially designed for MassARRAY, the assays could not be optimized for 14 SNPs, including all six SNPs from CYP2D6. Among the remaining 275 SNPs in 43 genes that were successfully genotyped, 122 SNPs and three genes had to be excluded for various reasons such as monoallelism, minor alleles having a frequency of less than 1% in our study populations, deviations from Hardy-Weinberg equilibrium with less than 5% significance, or revisions of the GenBank database relocating some SNPs to other genes.

Allele frequencies of the remaining 153 SNPs in 40 genes shown in Table 1 were calculated by combining the allele frequencies from the five geographical regions. There was no evidence of significant differences in allele frequencies among the five geographical regions (Fisher's exact test with Bonferroni correction), although such a difference might be detected by an analysis with a larger sample size. These SNPs showed Hardy-Weinberg equilibrium for two alleles, and the minor allele frequencies were more than 1%. The allele frequencies of 46 of the SNPs analyzed in this study were also reported in the JSNP project (Hirakawa et al. 2002), and significant differences were reported for only four SNPs ($p < 0.05$, Fisher's exact test). In the JSNP study, allele frequencies were determined for 752 unrelated Japanese volunteer subjects; further demographic details are not available.

Among the 50 polymorphisms in 241 noncancer Japanese outpatients at the Aichi Cancer Center Hospital reported by Hamajima et al. (2002), we report the following minor allele frequencies for the ten SNPs that overlapped with those in our study: Glu487Lys in ALDH2 (rs671), 0.278; Val158Met in COMT (rs4680), 0.346; T-34C in CYP17 (rs743572), 0.435; Ser326Cys in OGG1 (rs1052133), 0.471; C-889T in IL1A (rs1800587), 0.085; C-31T and C-511T in IL1B (rs1143627 and rs16944), 0.450 and 0.441, respectively; Ala223Val and Glu430Ala in MTHFR (rs1801133 and rs1801131), 0.405 and 0.193, respectively; and Pro187Ser in NQO1 (rs1800566), 0.421. Only the allele frequency for ALDH2 ($p < 0.05$, Fisher's exact test) was significantly different from the frequencies observed in the present study. As this study was a hospital-based study, it might not be comparable with our population-based study using a random sample; further studies will be needed to confirm the reason for the difference in allele frequencies of ALDH2.

The enrollment of representative subjects is essential for studying allele frequencies in a population of an area. However, there are few studies using a strict random sampling method in Japanese populations. In a study using a random sample of 445 Japanese rural residents in Hyogo Prefecture, Lwin et al. (2002) reported that the minor allele frequency of C677T in MTHFR (rs1801133) gene was 0.40, a frequency not significantly different from our study.

Here we report allele frequencies of SNPs in genes that may modify the dose-dependency or effects of exposure to cancer risk factors in a population-based study with random sampling in Japan. The primary aim of this study is to offer basic information on the genetic background of the Japanese population, which is highly useful for designing genome-based epidemiological studies, especially in the field of cancer research. However, the data may also be utilized as an alternative control population in exploratory genetic association studies for screening disease-related genes. More in-depth analyses are underway such as assessing the possible differences in allele/genotype frequencies among the populations from the five prefectures, assessing population stratification in Japan, and assessing gene-life style interactions.

Acknowledgements This study was supported mainly by the Program for Promotion of Fundamental Studies in Health Sciences of the Organization for Pharmaceutical Safety and Research of Japan; the Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology; and the Grant-in-Aid for Cancer Research from the Ministry of Health, Labour and Welfare.

References

Hamajima N, Saito T, Matsuo K, Suzuki T, Nakamura T, Matsuura A, Okuma K, Tajima K (2002) Genotype frequencies of 50 polymorphisms for 241 Japanese non-cancer patients. J Epidemiol 12:229–236

Hirakawa M, Tanaka T, Hashimoto Y, Kuroda M, Takagi T, Nakamura Y (2002) JSNP: a database of common gene variations in the Japanese population. Nucleic Acids Res 30:158–162

Lwin H, Yokoyama T, Date C, Yoshiike N, Kokubo Y, Tanaka H (2002) Are the associations between life-style related factors and plasma total homocysteine concentration different according to polymorphism of 5,10-methylenetetrahydrofolate reductase gene (C677T MTHFR)? A cross-sectional study in a Japanese rural population. J Epidemiol 12:126–135

Ross P, Hall L, Smirnov I, Haff L (1998) High level multiplex genotyping by MALDI-TOF mass spectrometry. Nat Biotechnol 16:1347–1351

Sugimura H, Wakai K, Genka K, Nagura K, Igarashi H, Nagayama K, Ohkawa A, Baba S, Morris BJ, Tsugane S, Ohno Y, Gao C, Li Z, Takezaki T, Tajima K, Iwamasa T (1998) Association of Ile462Val (Exon 7) polymorphism of cytochrome P450 IA1 with lung cancer in the Asian population: further evidence from a case-control study in Okinawa. Cancer Epidemiol Biomarkers Prev 7:413–417

Tsugane S, Gey F, Ichinowatari Y, Miyajima Y, Ishibashi T, Matsuishima S, Hirota Y, Inami T, Yamaguchi M, Karita K, Kabuto M, Takeshima Y, Todoriki H, Tsuda M, Akabane M, Furuichi Y, Hamada G, Watanabe S (1992a) Cross-sectional epidemiologic study for assessing cancer risks at the population level I. Study design and participation rate. J Epidemiol 2:75–81

Tsugane S, Gey F, Ichinowatari Y, Miyajima Y, Ishibashi T, Matsuishima S, Hirota Y, Inami T, Yamaguchi M, Karita K, Kabuto M, Takeshima Y, Todoriki H, Tsuda M, Akabane M, Furuichi Y, Hamada G, Watanabe S (1992b) Cross-sectional epidemiologic study for assessing cancer risks at the population level II. Baseline data and correlation analysis. J Epidemiol 2:83–89