SUPPLEMENTARY INFORMATION

Identification of Ethnically Specific Genetic Variations in Pan-Asian Ethnos

Jin Ok Yang[1]*, Sohyun Hwang[1], Woo-Yeon Kim[2], Seong-Jin Park[1], Sang Cheol Kim[3], Kiejung Park[1], Byungwook Lee[1], The HUGO Pan-Asian SNP Consortium[†]**

[1]Korean BioInformation Center (KOBIC), Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 305-806, Korea, [2]Bioinformatics Team, CSP R&D Center, Samsung SDS, Seoul 135-918, Korea, [3]Samsung Genome Institute, Samsung Medical Center, Seoul 135-710, Korea

http://www.genominfo.org/src/sm/gni-12-42-s001.pdf.

The participants of the HUGO Pan-Asian SNP Consortium are arranged by surname alphabetically in the following.

Mahmood Ameen Abdulla[1], Ikhlak Ahmed[2], Anunchai Assawamakin[3,4], Jong Bhak[5], Samir K. Brahmachari[2], Gayvelline C. Calacal[6], Amit Chaurasia[2], Chien-Hsiun Chen[7], Jieming Chen[8], Yuan-Tsong Chen[7], Jiayou Chu[9], Eva Maria C. Cutiongco-de la Paz[10], Maria Corazon A. De Ungria[6], Frederick C. Delfin[6], Juli Edo[1], Suthat Fuchareon[3], Ho Ghang[5], Takashi Gojobori[11,12], Junsong Han[13], Sheng-Feng Ho[7], Boon Peng Hoh[14], Wei Huang[15], Hidetoshi Inoko[16], Pankaj Jha[2], Timothy A. Jinam[1], Li Jin[17,18], Jongsun Jung[19], Daoroong Kangwanpong[20], Jatupol Kampuansai[20], Giulia C. Kennedy[21,22], Preeti Khurana[23], Hyung-Lae Kim[19], Kwangjoong Kim[19], Sangsoo Kim[24], Woo-Yeon Kim[5], Kuchan Kimm[25], Ryosuke Kimura[26], Tomohiro Koike[11], Supasak Kulawonganunchai[4], Vikrant Kumar[8], Poh San Lai[27,28], Jong-Young Lee[19], Sunghoon Lee[5], Edison T. Liu[8], Partha P. Majumder[29], Kiran Kumar Mandapati[23], Sangkot Marzuki[30], Wayne Mitchell[31,32], Mitali Mukerji[2], Kenji Naritomi[33], Chumpol Ngamphiw[4], Norio Niikawa[34], Nao Nishida[26], Bermseok Oh[19], Sangho Oh[5], Jun Ohashi[26], Akira Oka[16], Rick Ong[8], Carmencita D. Padilla[10], Prasit Palittapongarnpim[35], Henry B. Perdigon[6], Maude Elvira Phipps[1,36], Eileen Png[8], Yoshiyuki Sakaki[37], Jazelyn M. Salvador[6], Yuliana Sandraling[30], Vinod Scaria[2], Mark Seielstad[8], Mohd Ros Sidek[14], Amit Sinha[2], Metawee Srikummool[19], Herawati Sudoyo[30], Sumio Sugano[38], Helena Suryadi[30], Yoshiyuki Suzuki[11], Kristina A. Tabbada[6], Adrian Tan[8], Katsushi Tokunaga[26], Sissades Tongsima[4], Lilian P. Villamor[6], Eric Wang[21,22], Ying Wang[15], Haifeng Wang[15], Jer-Yuarn Wu[7], Huasheng Xiao[13], Shuhua Xu[18], Jin Ok Yang[5], Yin Yao Shugart[39], Hyang-Sook Yoo[5], Wentao Yuan[15], Guoping Zhao[15], Bin Alwi Zilfalil[14], Indian Genome Variation Consortium[2]

[1]Department of Molecular Medicine, Faculty of Medicine, and the Department of Anthropology, Faculty of Arts and Social Sciences, University of Malaya, Kuala Lumpur 50603, Malaysia, [2]Institute of Genomics and Integrative Biology, Council for Scientific and Industrial Research, Delhi 110007, India, [3]Mahidol University, Salaya Campus, Nakornpathom 73170, Thailand, [4]Biostatistics and Informatics Laboratory, Genome Institute, National Center for Genetic Engineering and Biotechnology, Thailand Science Park, Pathumtani 12120, Thailand, [5]Korean BioInformation Center (KOBIC), Korea Research Institute of Bioscience and Biotechnology (KRIBB), Deajeon 305-806, Korea, [6]DNA Analysis Laboratory, Natural Sciences Research Institute, University of the Philippines, Quezon City 1101, Philippines, [7]Institute of Biomedical Sciences, Academia Sinica, Taipei City 115, Taiwan, [8]Genome Institute of Singapore, Singapore 138672, Singapore, [9]Institute of Medical Biology, Chinese Academy of Medical Science, Kunming 650107, China, [10]Institute of Human Genetics, National Institutes of Health, University of the Philippines Manila, Ermita Manila 1000, Philippines, [11]Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Research Organization of Information and Systems, Mishima 411-8540, Japan, [12]Biomedicinal Information Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo 135-0064, Japan, [13]National Engineering Center for Biochip at Shanghai, Shanghai 201203, China, [14]Human Genome Center, School of Medical Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Malaysia, [15]MOST-Shanghai Laboratory of Disease and Health Genomics, Chinese National Human Genome Center Shanghai, Shanghai 201203, China, [16]Department of Molecular Life Science Division of Molecular Medical Science and Molecular Medicine, Tokai University School of Medicine, Isehara A259-1193, Japan, [17]State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life

Sciences, Fudan University, Shanghai 200433, China, [18]Chinese Academy of Sciences-Max Planck Society Partner Institute for Computational Biology, Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China, [19]Korea National Institute of Health, Seoul, 122-701, Korea, [20]Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50202, Thailand, [21]Genomics Collaborations, Affymetrix, Santa Clara, CA 95051, USA, [22]Veracyte, South San Francisco, CA 94080, USA, [23]The Centre for Genomic Applications (an IGIB-IMM Collaboration), New Delhi 110020, India, [24]Soongsil University, Seoul 156-743, Korea, [25]Eulji University College of Medicine, Daejeon 301-832, Korea, [26]Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo 113-0033, Japan, [27]Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, National University Hospital, Singapore 119074, Singapore, [28]Population Genetics Lab, Defence Medical and Environmental Research Institute, DSO National Laboratories, Singapore 117510, Singapore, [29]Indian Statistical Institute (Kolkata), Kolkata 700108, India, [30]Eijkman Institute for Molecular Biology, Jakarta 10430, Indonesia, [31]Informatics Experimental Therapeutic Centre, Singapore 138669, Singapore, [32]Division of Information Sciences, School of Computer Engineering, Nanyang Technological University, Singapore 639798, Singapore, [33]Department of Medical Genetics, University of the Ryukyus Faculty of Medicine, Nishihara 903-0215, Japan, [34]Research Institute of Personalized Health Sciences, Health Sciences University of Hokkaido, Tobetsu 061-0293, Japan, [35]National Science and Technology Development Agency, Pathumtani 12120, Thailand, [36]Monash University (Sunway Campus), Jalan Lagoon Selatan, 46150 Bandar Sunway, Selangor, Malaysia, [37]RIKEN Genomic Sciences Center, Yokohama 230-0045, Japan, [38]Laboratory of Functional Genomics, Department of Medical Genome Sciences Graduate School of Frontier Sciences, University of Tokyo (Shirokanedai Laboratory), Tokyo 108-8639, Japan, [39]Genomic Research Branch, National Institute of Mental Health, National Institutes of Health, 6001 Executive Boulevard, Bethesda, MD 20892 USA.

Supplementary Table 1. Ethnic groups list in Pan Asia (assign Ethnic group as like "two country-two ethnic group") and HapMap populations

Abbreviation of ethnic group	Ethnicity	Abbreviation of ethnic group	Ethnicity	Abbreviation of ethnic group	Ethnicity
AX-AI	Karitiana	ID-SU	Sunda	PI-MA	Minanubu
AX-AM	Ami	ID-TB	Batak Toba	PI-MW	Mamanwa
AX-AT	Atayal	ID-TR	Toraja	PI-UB	Filipino
AX-ME	Melanesians	IN-DR	Proto-Austroloids	PI-UI	Filipino
CEU	European	IN-EL	Caucasoids	PI-UN	Filipino
CHB	Han	IN-IL	Caucasoids	SG-CH	Chinese
CN-CC	Zhuang	IN-NI	Mongoloid features	SG-ID	Indian
CN-GA	Han	IN-NL	Caucasoids	SG-ML	Malay
CN-HM	Hmong	IN-SP	Caucasoids	TH-HM	Hmong
CN-JI	Jiamao	IN-TB	Mongoloid features	TH-KA	Karen
CN-JN	Jinuo	IN-WI	Caucasoids	TH-LW	Lawa
CN-SH	Han	IN-WL	Caucasoids	TH-MA	Mlabri
CN-UG	Uyghur	JP-ML	Japanese	TH-MO	Mon
CN-WA	Wa	JP-RK	Ryukyuan	TH-PL	Paluang
ID-AL	Alorese	JPT	Japanese	TH-PP	Plang
ID-DY	Dayak	KR-KR	Koreans	TH-TK	Tai Khuen
ID-JA	Javanese	MY-BD	Bidayuh	TH-TL	Tai Lue
ID-JV	Javanese	MY-JH	Negrito	TH-TN	H'tin
ID-KR	Batak Karo	MY-KN	Malay	TH-TU	Tai Yuan
ID-LA	Lamaholot	MY-KS	Negrito	TH-TY	Tai Yong
ID-LE	Lembata	MY-MN	Malay	TH-YA	Yao
ID-ML	Malay	MY-TM	Proto-Malay	TW-HA	Chinese
ID-MT	Mentawai	PI-AE	Ayta	TW-HB	Chinese
ID-RA	Manggarai	PI-AG	Agta	YRI	Yoruba
ID-SB	Kambera	PI-AT	Ati		
ID-SO	Manggarai	PI-IR	Iraya		

AX, Affymetrix Inc. (not country); CN, China; ID, Indonesia; IN, India; JP, Japan; KR, Korea; MY, Malaysia; PI, Philippine; SG, Singapore; TH, Thailand; TW, Taiwan.

The samples from HapMap were expressed as three codes: CEU, European sampled in America; CHB, Chinese sampled in US; JPT, Japanese sampled in Japan; YRI, Yoruba sampled in Africa.

Supplementary Table 2. ESNP- and ECNP-related genes list

ESNP-related genes	Description	Location
PCCA	Propionyl CoA carboxylase, alpha polypeptide	Cytoplasm
IGHMBP2	Immunoglobulin mu binding protein 2	Nucleus
RNF43	Ring finger protein 43	Plasma membrane
ALAD	Aminolevulinate dehydratase	Cytoplasm
CS	Citrate synthase	Cytoplasm
CARS2	Cysteinyl-tRNA synthetase 2, mitochondrial (putative)	Cytoplasm
TPO	Thyroid peroxidase	Plasma membrane
RGS7	Regulator of G-protein signaling 7	Cytoplasm
PLA2G4C	Phospholipase A2, group IVC (cytosolic, calcium-independent)	Plasma membrane
LYPLA1	Lysophospholipase I	Cytoplasm
BRIP1	BRCA1 interacting protein C-terminal helicase 1	Nucleus
PJA2	Praja ring finger 2, E3 ubiquitin protein ligase	Cytoplasm
**SEP9*	Septin 9	Cytoplasm
GBE1	Glucan (1,4-alpha-), branching enzyme 1	Cytoplasm
ZNRF1	Zinc and ring finger 1, E3 ubiquitin protein ligase	Cytoplasm
TP53I3	Tumor protein p53 inducible protein 3	Other
UBR2	Ubiquitin protein ligase E3 component n-recognin 2	Nucleus
TDP1	Tyrosyl-DNA phosphodiesterase 1	Nucleus
EMR3	Egf-like module containing, mucin-like, hormone receptor-like 3	Plasma membrane
LPHN3	Latrophilin 3	Plasma membrane
LTBP4	Latent transforming growth factor beta binding protein 4	Extracellular space
KCNMA1	Potassium large conductance calcium-activated channel, subfamily M, alpha member 1	Plasma membrane
PRKCE	Protein kinase C, epsilon	Cytoplasm
GOLGA5	Golgin A5	Cytoplasm
ITPK1	Inositol-tetrakisphosphate 1-kinase	Cytoplasm
ITPKB	Inositol-trisphosphate 3-kinase B	Cytoplasm
MYLK3	Myosin light chain kinase 3	Cytoplasm
GRK5	G protein-coupled receptor kinase 5	Plasma membrane
IPPK	Inositol 1,3,4,5,6-pentakisphosphate 2-kinase	Cytoplasm
PRKD2	Protein kinase D2	Cytoplasm
GALK2	Galactokinase 2	Cytoplasm
PKIA	Protein kinase (cAMP-dependent, catalytic) inhibitor alpha	Cytoplasm
GRAP2	GRB2-related adaptor protein 2	Cytoplasm

Supplementary Table 2. Continued.

ESNP-related genes	Description	Location
FAM107B	Family with sequence similarity 107, member B	Other
LRRC20	Leucine rich repeat containing 20	Other
SMEK1	SMEK homolog 1, suppressor of mek1 (Dictyostelium)	Plasma membrane
FBXW4	F-box and WD repeat domain containing 4	Other
RALGPS1	Ral GEF with PH domain and SH3 binding motif 1	Cytoplasm
AHRR	Aryl-hydrocarbon receptor repressor	Nucleus
SLC17A9	Solute carrier family 17, member 9	Other
TMEM143	Transmembrane protein 143	Cytoplasm
CAMTA1	Calmodulin binding transcription activator 1	Other
CUL9	Cullin 9	Cytoplasm
MSL1	Male-specific lethal 1 homolog (Drosophila)	Nucleus
FAM118A	Family with sequence similarity 118, member A	Other
TMEM170A	Transmembrane protein 170A	Other
TUBGCP3	Tubulin, gamma complex associated protein 3	Cytoplasm
NIPSNAP3A	Nipsnap homolog 3A (C. elegans)	Cytoplasm
CARD8	Caspase recruitment domain family, member 8	Nucleus
LOC100506325	Uncharacterized LOC100506325	Other
PROS1	Protein S (alpha)	Extracellular space
SETBP1	SET binding protein 1	Nucleus
ECM2	Extracellular matrix protein 2, female organ and adipocyte specific	Extracellular space
FAT4	FAT atypical cadherin 4	Other
DCUN1D2	DCN1, defective in cullin neddylation 1, domain containing 2	Other
COL23A1	Collagen, type XXIII, alpha 1	Plasma membrane
C17orf67	Chromosome 17 open reading frame 67	Other
COL26A1	Collagen, type XXVI, alpha 1	Extracellular space
FAM196A	Family with sequence similarity 196, member A	Other
URGCP	Upregulator of cell proliferation	Cytoplasm
SPON1	Spondin 1, extracellular matrix protein	Extracellular space
ANK2	Ankyrin 2, neuronal	Plasma membrane
RFC2	Replication factor C (activator 1) 2, 40kDa	Nucleus
ABLIM2	Actin binding LIM protein family, member 2	Cytoplasm
KIAA0930	KIAA0930	Other

Supplementary Table 2. Continued.

ESNP-related genes	Description	Location
ELN	Elastin	Extracellular space
SUMO3	Small ubiquitin-like modifier 3	Nucleus
ARFGEF1	ADP-ribosylation factor guanine nucleotide-exchange factor 1 (brefeldin A-inhibited)	Cytoplasm
FAM13B	Family with sequence similarity 13, member B	Cytoplasm
SPACA7	Sperm acrosome associated 7	Cytoplasm
MEI1	Meiosis inhibitor 1	Other
ZC3H4	Zinc finger CCCH-type containing 4	Other
FBXL20	F-box and leucine-rich repeat protein 20	Cytoplasm
MBP	Myelin basic protein	Extracellular space
KIAA1244	KIAA1244	Cytoplasm
COLGALT1	Collagen beta(1-O)galactosyltransferase 1	Cytoplasm
ZBTB46	Xinc finger and BTB domain containing 46	Nucleus
KDM4B	Lysine (K)-specific demethylase 4B	Other
WDR89	WD repeat domain 89	Other
SH3BP2	SH3-domain binding protein 2	Cytoplasm
GTDC1	Glycosyltransferase-like domain containing 1	Other
MTMR11	Myotubularin related protein 11	Other
HCG18	HLA complex group 18 (non-protein coding)	Other
CHGA	Chromogranin A (parathyroid secretory protein 1)	Cytoplasm
GLG1	Golgi glycoprotein 1	Cytoplasm
LOC285972	Uncharacterized LOC285972	Other
EVI5	Ecotropic viral integration site 5	Other
LOC100130172	Uncharacterized LOC100130172	Other
DIDO1	Death inducer-obliterator 1	Nucleus
PHF20	PHD finger protein 20	Nucleus
PDZD2	PDZ domain containing 2	Plasma membrane
CEP192	Centrosomal protein 192kDa	Cytoplasm
TRMT44	tRNA methyltransferase 44 homolog (S. cerevisiae)	Other
RIN3	Ras and Rab interactor 3	Cytoplasm

Supplementary Table 2. Continued.

ESNP-related genes	Description	Location
NF1	Neurofibromin 1	Cytoplasm
CENPM	Centromere protein M	Cytoplasm
CPLX2	Complexin 2	Cytoplasm
ERC1	ELKS/RAB6-interacting/CAST family member 1	Cytoplasm
VWA3B	Von Willebrand factor A domain containing 3B	Other
LRRC36	Leucine rich repeat containing 36	Other
UTS2	Urotensin 2	Extracellular space
COL28A1	Collagen, type XXVIII, alpha 1	Extracellular space
TNRC6C	Trinucleotide repeat containing 6C	Cytoplasm
EFCAB11	EF-hand calcium binding domain 11	Other
CCNY	Cyclin Y	Nucleus
VOPP1	Vesicular, overexpressed in cancer, prosurvival protein 1	Nucleus
NLRC3	NLR family, CARD domain containing 3	Cytoplasm
FCHO1	FCH domain only 1	Plasma membrane
LOC100506172	Uncharacterized LOC100506172	Other
TUBGCP6	Tubulin, gamma complex associated protein 6	Cytoplasm
C1orf54	Chromosome 1 open reading frame 54	Other
RASGRF1	Ras protein-specific guanine nucleotide-releasing factor 1	Cytoplasm
MADD	MAP-kinase activating death domain	Cytoplasm
ZNF484	Zinc finger protein 484	Nucleus
ACTR3C	ARP3 actin-related protein 3 homolog C (yeast)	Other
PRC1	Protein regulator of cytokinesis 1	Nucleus
TBC1D22A	TBC1 domain family, member 22A	Other
OTUD7B	OTU domain containing 7B	Cytoplasm
CASP9	Caspase 9, apoptosis-related cysteine peptidase	Cytoplasm
AEBP1	AE binding protein 1	Nucleus
PTPRA	Protein tyrosine phosphatase, receptor type, A	Plasma membrane
HOXD3	Homeobox D3	Nucleus
NPAS2	Neuronal PAS domain protein 2	Nucleus

Supplementary Table 2. Continued.

ESNP-related genes	Description	Location
ZNF236	Zinc finger protein 236	Nucleus
TRERF1	Transcriptional regulating factor 1	Nucleus
VAV2	Vav 2 guanine nucleotide exchange factor	Cytoplasm
ZNF174	Zinc finger protein 174	Nucleus
GATAD2B	GATA zinc finger domain containing 2B	Nucleus
UBN1	Ubinuclein 1	Nucleus
RUNX3	Runt-related transcription factor 3	Nucleus
PRRX2	Paired related homeobox 2	Nucleus
ELF1	E74-like factor 1 (ets domain transcription factor)	Nucleus
MXI1	MAX interactor 1, dimerization protein	Nucleus
CERS4	Ceramide synthase 4	Cytoplasm
KLF12	Kruppel-like factor 12	Nucleus
EIF4E2	Eukaryotic translation initiation factor 4E family member 2	Cytoplasm
EIF4H	Eukaryotic translation initiation factor 4H	Cytoplasm
UTRN	Utrophin	Plasma membrane
ROBO1	Roundabout, axon guidance receptor, homolog 1 (Drosophila)	Plasma membrane
ITGB3	Integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)	Plasma membrane
TNFRSF9	Tumor necrosis factor receptor superfamily, member 9	Plasma membrane
ABCA1	ATP-binding cassette, sub-family A (ABC1), member 1	Plasma membrane
SMC1B	Structural maintenance of chromosomes 1B	Nucleus
XPO6	Exportin 6	Cytoplasm
NUP50	Nucleoporin 50kDa	Nucleus
MIP	Major intrinsic protein of lens fiber	Plasma membrane
STX11	Syntaxin 11	Plasma membrane
TMCO3	Transmembrane and coiled-coil domains 3	Other
VPS45	Vacuolar protein sorting 45 homolog (S. cerevisiae)	Cytoplasm
EXOC3	Exocyst complex component 3	Plasma membrane
ECNP-related genes	Description	Location
APOBEC2	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 2	Other
NDUFA9	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 9, 39kDa	Cytoplasm
OARD1	O-acyl-ADP-ribose deacylase 1	Other
TOP1	Topoisomerase (DNA) I	Nucleus

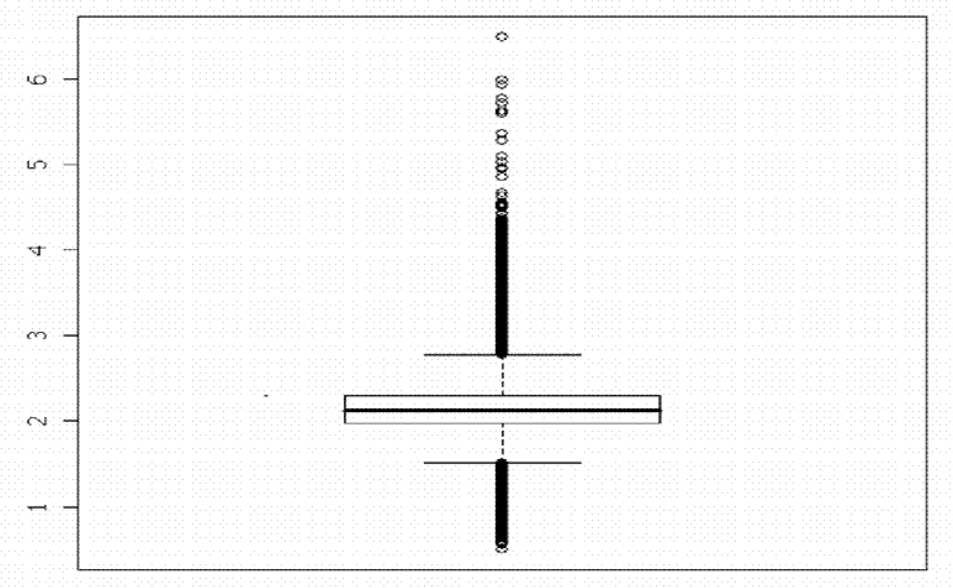
Supplementary Table 2. Continued.

ECNP-related genes	Description	Location
UAP1	UDP-N-acteylglucosamine pyrophosphorylase 1	Nucleus
GABRA5	Gamma-aminobutyric acid (GABA) A receptor, alpha 5	Plasma membrane
GABRG3	Gamma-aminobutyric acid (GABA) A receptor, gamma 3	Plasma membrane
DDR2	Discoidin domain receptor tyrosine kinase 2	Plasma membrane
ALCAM	Activated leukocyte cell adhesion molecule	Plasma membrane
ASTN2	Astrotactin 2	Cytoplasm
C12orf4	Chromosome 12 open reading frame 4	Other
C1QTNF7	C1q and tumor necrosis factor related protein 7	Extracellular space
C3orf20	Chromosome 3 open reading frame 20	Cytoplasm
CCDC174	Coiled-coil domain containing 174	Other
COL6A4P1	Collagen, type VI, alpha 4 pseudogene 1	Other
CPEB2	Cytoplasmic polyadenylation element binding protein 2	Cytoplasm
DAB1	Dab, reelin signal transducer, homolog 1 (Drosophila)	Cytoplasm
DLGAP3	Discs, large (Drosophila) homolog-associated protein 3	Cytoplasm
FGD5	FYVE, RhoGEF and PH domain containing 5	Cytoplasm
FLJ43663	Uncharacterized LOC378805	Other
FNDC1	Fibronectin type III domain containing 1	Plasma membrane
GTF2H5	General transcription factor IIH, polypeptide 5	Nucleus
KIAA0319L	KIAA0319-like	Cytoplasm
LINC00111	Long intergenic non-protein coding RNA 111	Other
LOC152742	Uncharacterized LOC152742	Other
LOC441009	Uncharacterized LOC441009	Other
MAG	Myelin associated glycoprotein	Plasma membrane
NUF2	NUF2, NDC80 kinetochore complex component	Nucleus
RAD51AP1	RAD51 associated protein 1	Nucleus
RBFOX1	RNA binding protein, fox-1 homolog (C. elegans) 1	Cytoplasm
RGS5	Regulator of G-protein signaling 5	Plasma membrane
SERAC1	Serine active site containing 1	Extracellular space
TAGAP	T-cell activation RhoGTPase activating protein	Cytoplasm
TCERG1L	Transcription elongation regulator 1-like	Other
TREML4	Triggering receptor expressed on myeloid cells-like 4	Other
TRMT44	tRNA methyltransferase 44 homolog (S. cerevisiae)	Other
ZFYVE20	Zinc finger, FYVE domain containing 20	Cytoplasm
ZMYM4	Zinc finger, MYM-type 4	Other

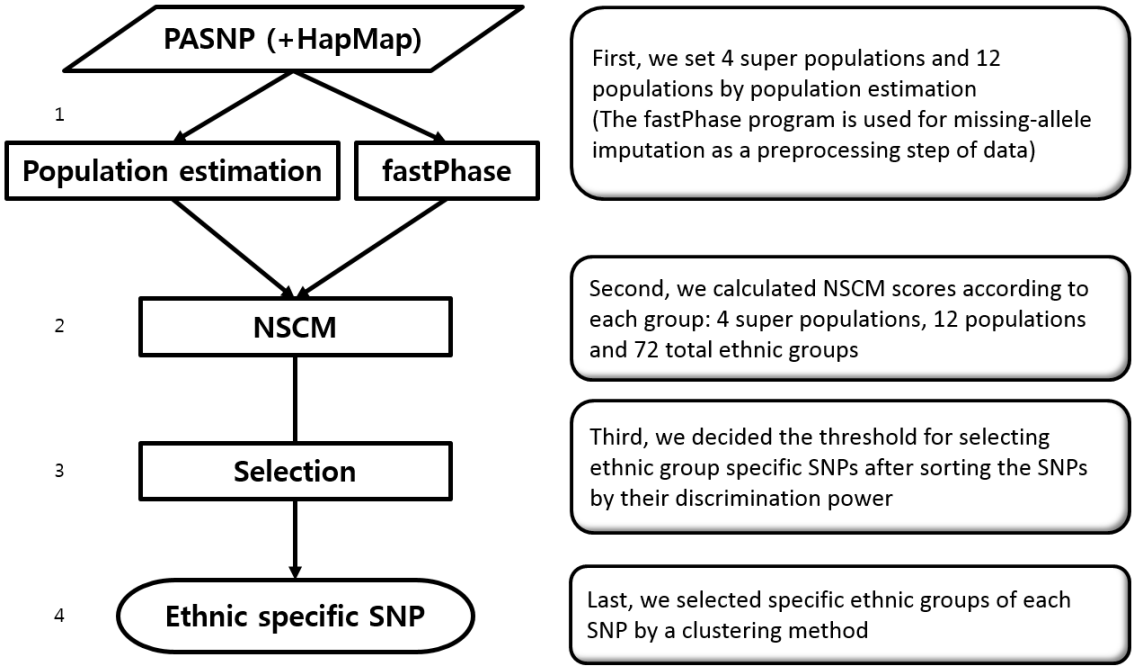
Supplementary Table 2. Continued.

ECNP-related genes	Description	Location
CAPN7	Calpain 7	Cytoplasm
PSMB2	Proteasome (prosome, macropain) subunit, beta type, 2	Cytoplasm
SYNJ2	Synaptojanin 2	Cytoplasm
ARID1B	AT rich interactive domain 1B (SWI1-like)	Nucleus
MYT1L	Myelin transcription factor 1-like	Nucleus
NFYA	Nuclear transcription factor Y, alpha	Nucleus
SPEN	Spen homolog, transcriptional regulator (Drosophila)	Nucleus
TULP4	Tubby like protein 4	Cytoplasm
NCR2	Natural cytotoxicity triggering receptor 2	Plasma membrane
TLR4	Toll-like receptor 4	Plasma membrane
OCA2	Oculocutaneous albinism II	Plasma membrane
SLC22A1	Solute carrier family 22 (organic cation transporter), member 1	Plasma membrane
SLC22A3	Solute carrier family 22 (extraneuronal monoamine transporter), member 3	Plasma membrane
SNX9	Sorting nexin 9	Cytoplasm

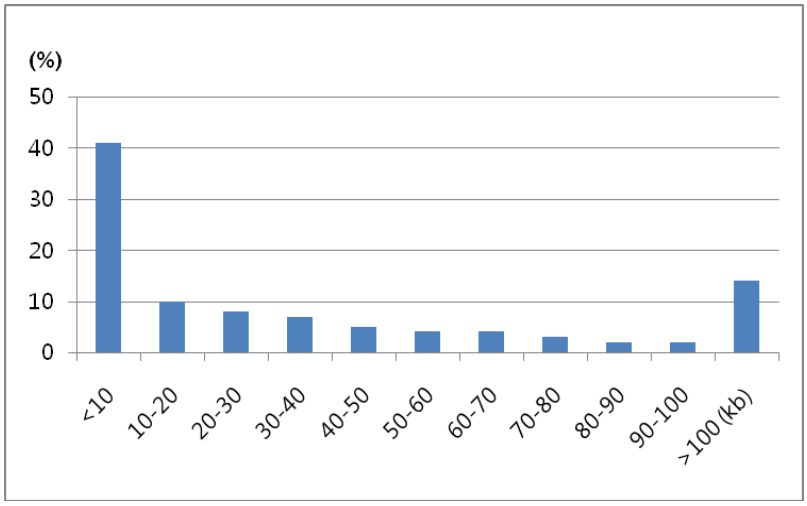
ESNP, ethnic variant single-nucleotide polymorphism; ECNP, ethnic variant copy number polymorphism.



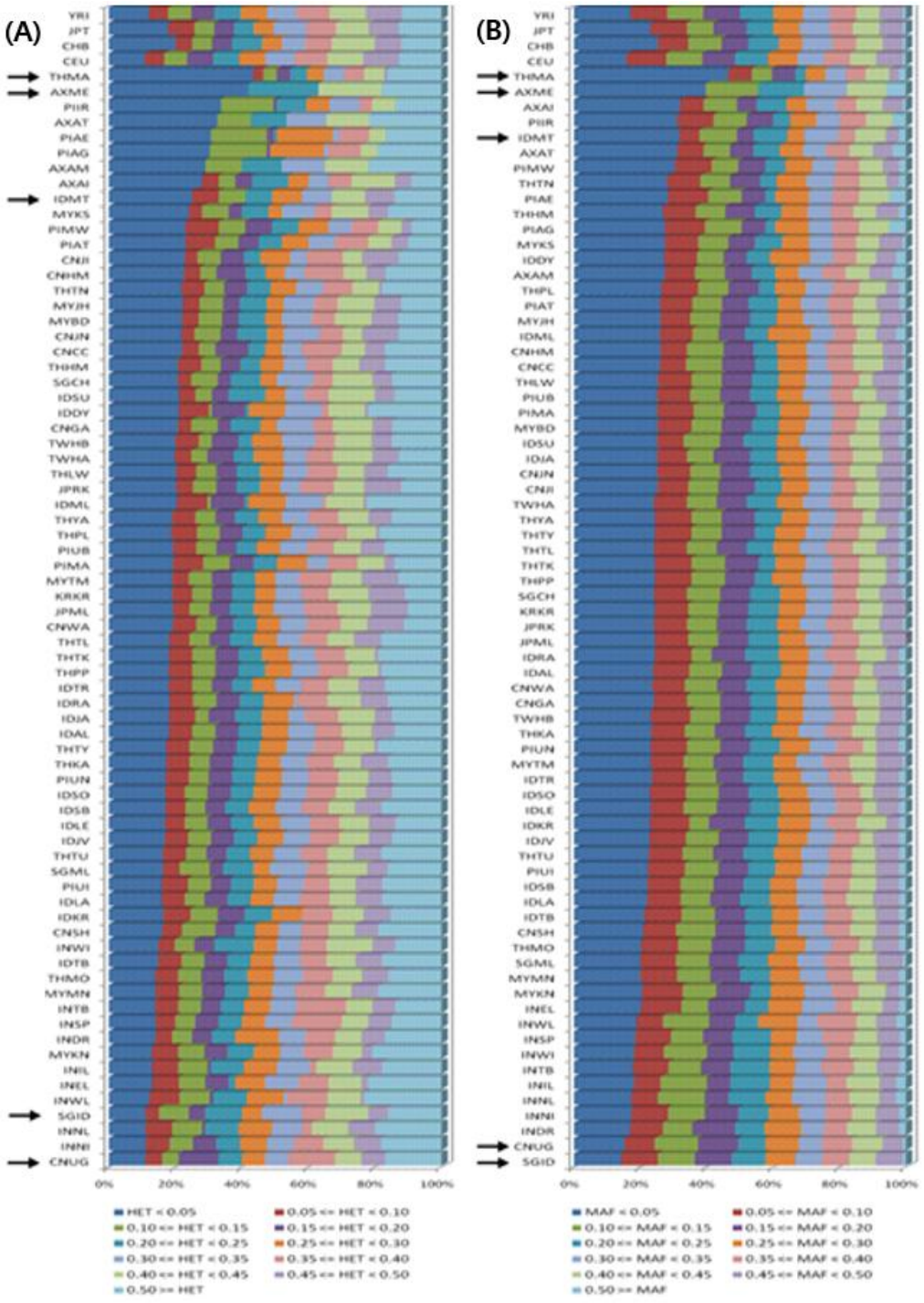
Supplementary Fig. 1. Distribution of Pan-Asian copy number values from Pan-Asian genotype profiling. It shows that discrete distribution of Pan-Asian single nucleotide polymorphism samples with lower boundary of 1.5065 and upper boundary of 2.7765.



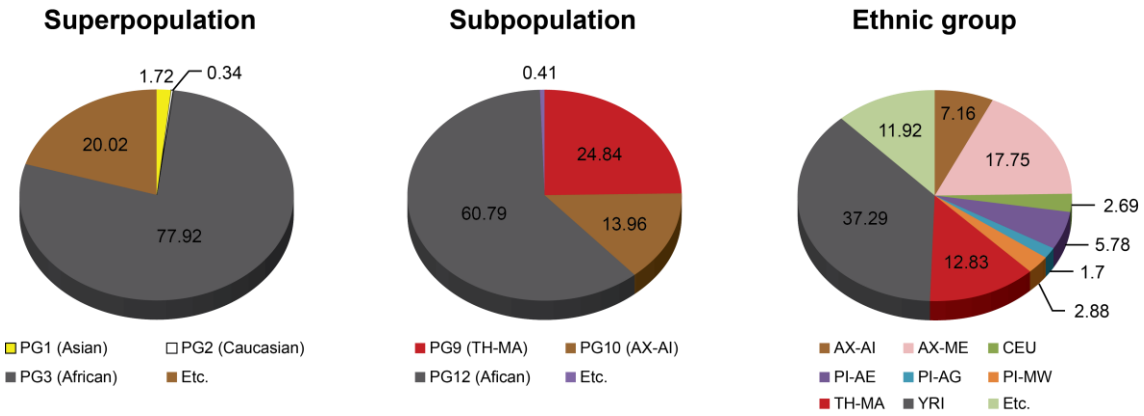
Supplementary Fig. 2. Process to select ethnic specific single nucleotide polymorphism (SNP). NSCM, nearest shrunken centroid method.



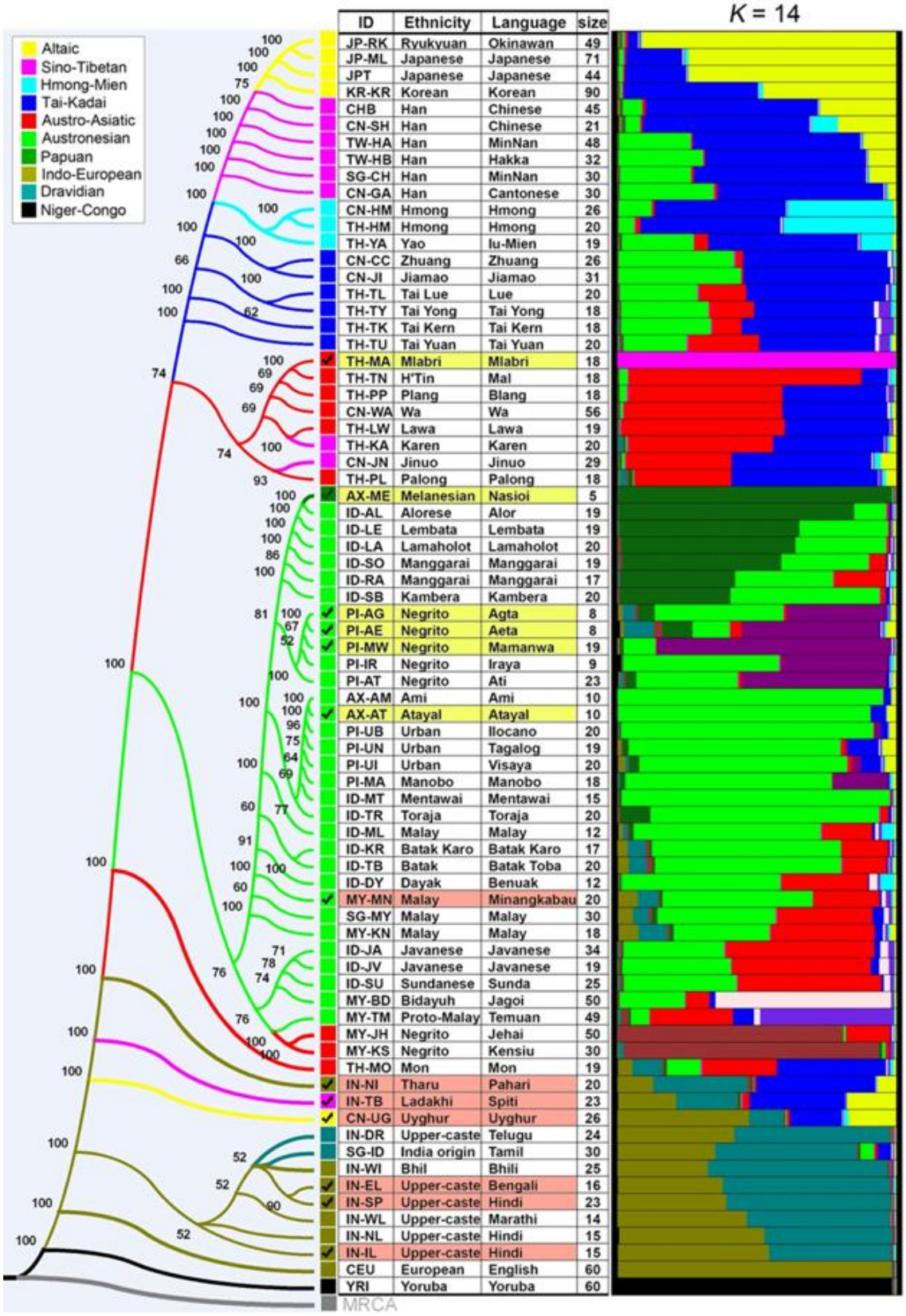
Supplementary Fig. 3. Inter-single nucleotide polymorphism (SNP) distance distribution. The X-axis represents the distance (kilo base pair) between SNPs and the Y-axis represents the proportion (%) of pan Asian SNPs.



Supplementary Fig. 4. Minor allele frequency (MAF) and heterozygosity (HET) distribution of single nucleotide polymorphism (SNP) in the Pan-Asian SNP data. X-axis is Pan-Asian ethnic groups, and Y-axis is SNP proportion (%) for MAF in each range. We examined MAF and HET across ethnic groups and 4 HapMap groups together. We assigned SNP proportion (%) in each range for MAF and HET rate of Pan-Asia ethnic groups and 4 HapMap groups as shown in (A) and (B).



Supplementary Fig. 5. The distribution of ethnic variant single nucleotide polymorphisms (ESNPs) across Pan-Asian and HapMap ethnic groups. We analyzed the ethnicity-specific single-nucleotide polymorphisms, including the four HapMap groups (CEU, CHB, JPT, and YRI), for the following groups: super-populations (Asians, Caucasoids, American Indians, and outliers [IN-NI, IN-TB, and CN-UG]); 12 populations; and 76 ethnic groups (Pan-Asian and four HapMap ethnic groups). AX-AI, Karitiana; AX-ME, Ami; CEU, European; PI-AE, Ayta; PI-AG, Ayta; PI-MW, Mamanwa; TH-MA, Mlabri; YRI, Yoruba; CHB, Han; JPT, Japanese; IN-NI, Mongoloid features; IN-TB, Mongoloid features; CN-UG, Uyghur.



ID	Ethnicity	Language	size
JP-RK	Ryukyuan	Okinawan	49
JP-ML	Japanese	Japanese	71
JPT	Japanese	Japanese	44
KR-KR	Korean	Korean	90
CHB	Han	Chinese	45
CN-SH	Han	Chinese	21
TW-HA	Han	MinNan	48
TW-HB	Han	Hakka	32
SG-CH	Han	MinNan	30
CN-GA	Han	Cantonese	30
CN-HM	Hmong	Hmong	26
TH-HM	Hmong	Hmong	20
TH-YA	Yao	Iu-Mien	19
CN-CC	Zhuang	Zhuang	26
CN-JI	Jiamao	Jiamao	31
TH-TL	Tai Lue	Lue	20
TH-TY	Tai Yong	Tai Yong	18
TH-TK	Tai Kern	Tai Kern	18
TH-TU	Tai Yuan	Tai Yuan	20
TH-MA	Mlabri	Mlabri	18
TH-TN	H'Tin	Mal	18
TH-PP	Plang	Blang	18
CN-WA	Wa	Wa	56
TH-LW	Lawa	Lawa	19
TH-KA	Karen	Karen	20
CN-JN	Jinuo	Jinuo	29
TH-PL	Palong	Palong	18
AX-ME	Melanesian	Nasioi	5
ID-AL	Alorese	Alor	19
ID-LE	Lembata	Lembata	19
ID-LA	Lamaholot	Lamaholot	20
ID-SO	Manggarai	Manggarai	19
ID-RA	Manggarai	Manggarai	17
ID-SB	Kambera	Kambera	20
PI-AG	Negrito	Agta	8
PI-AE	Negrito	Aeta	8
PI-MW	Negrito	Mamanwa	19
PI-IR	Negrito	Iraya	9
PI-AT	Negrito	Ati	23
AX-AM	Ami	Ami	10
AX-AT	Atayal	Atayal	10
PI-UB	Urban	Ilocano	20
PI-UN	Urban	Tagalog	19
PI-UI	Urban	Visaya	20
PI-MA	Manobo	Manobo	18
ID-MT	Mentawai	Mentawai	15
ID-TR	Toraja	Toraja	20
ID-ML	Malay	Malay	12
ID-KR	Batak Karo	Batak Karo	17
ID-TB	Batak	Batak Toba	20
ID-DY	Dayak	Benuak	12
MY-MN	Malay	Minangkabau	20
SG-MY	Malay	Malay	30
MY-KN	Malay	Malay	18
ID-JA	Javanese	Javanese	34
ID-JV	Javanese	Javanese	19
ID-SU	Sundanese	Sunda	25
MY-BD	Bidayuh	Jagoi	50
MY-TM	Proto-Malay	Temuan	49
MY-JH	Negrito	Jehai	50
MY-KS	Negrito	Kensiu	30
TH-MO	Mon	Mon	19
IN-NI	Tharu	Pahari	20
IN-TB	Ladakhi	Spiti	23
CN-UG	Uyghur	Uyghur	26
IN-DR	Upper-caste	Telugu	24
SG-ID	India origin	Tamil	30
IN-WI	Bhil	Bhili	25
IN-EL	Upper-caste	Bengali	16
IN-SP	Upper-caste	Hindi	23
IN-WL	Upper-caste	Marathi	14
IN-NL	Upper-caste	Hindi	15
IN-IL	Upper-caste	Hindi	15
CEU	European	English	60
YRI	Yoruba	Yoruba	60

Supplementary Fig. 6. Representative ethnic groups having ethnically specific single nucleotide

polymorphismss and copy number polymorphisms on population structures. Yellow-colored row indicates the Pan-Asian ethnic groups having highly portion of ethnic variant single nucleotide polymorphisms and red-colored row indicates the Pan-Asian ethnic groups having highly portion of ethnic variant copy number polymorphisms. We marked the Pan-Asian ethnic groups based on a maximum-likelihood tree of populations. Abbreviations are explained in Supplementary Table 1.