Supplementary Information for: Fosmid-based whole genome haplotyping of a HapMap trio child: Evaluation of Single Individual Haplotyping techniques

Jorge Duitama [1,2,*,‡], Gayle McEwen [1,*], Thomas Huebsch [1], Stefanie Palczewski [1], Sabrina Schulz [1], Kevin Verstrepen [2], Eun-Kyung Suk [1,†], and Margret R. Hoehe [1,†,‡]

[1]Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195, Berlin, Germany
[2]VIB Laboratory of Systems Biology & Laboratory for Genetics and Genomics, Center of Microbial and Plant Genetics, K.U.Leuven, Gaston Geenslaan 1, B-3001 Leuven (Heverlee), Belgium

[*]The authors contributed equally to this work
[†]Co-last authors
[‡]To whom correspondence should be addressed. Dr. Margret R. Hoehe, Email: hoehe@molgen.mpg.de, Tel: + 49 30 8413 1468; Fax: + 49 30 8413 1462. Correspondence can also be addressed to Jorge Duitama, Email Jorge.DuitamaCastellanos@biw.vib-kuleuven.be, Tel: + 32 1675 1402; Fax: + 32 1675 1391

© 2008 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

chr	Start	End	Number of additional SNPS phased	Disease associated Gene(GAD)	Disease relevance
chr7	132588363	133401053	303	EXOC4	Amyotrophic Lateral Sclerosis Diabetes Mellitus, Type 2
chr6	31344508	31432914	285	HLA-B/HLA-C	Lymphoma, Non-Hodgkin Rheumatoid spondylitis
chr6	161689662	163068790	255	PARK2	Parkinson Disease, Type 2 Diabetes
chr2	29269144	29997936	222	ALK	Multiple sclerosis
chr12	2032725	2677376	187	CACNA1C	Bipolar disorder, Narcolepsy
chr2	211948687	213111597	184	ERBB4	Lung cancer, Schizophrenia
chr10	67349937	69125933	176	CTNNA3	Type 2 Diabetes, Multiple Sclerosis
chr6	29903803	30085613	165	HLA	Alopecia areata, Leukemia
chr6	29902723	30085613	165	HLA-A /HLA-G	Celiac Disease, Chorioamnionitis
chr17	28364219	29507938	153	ACCN1	Narcolepsy, Tobacco Use Disorder
chr3	59710076	61212164	148	FHIT	Lymphatic Metastasis Type 2 Diabetes
chr15	83724875	84093590	145	AKAP13	Tobacco Use Disorder Graft vs Host Disease
chr10	83625077	84735341	145	NRG3	Chronic renal failure Schizophrenia
chr6	163068154	163656514	134	PACRG	Acquired Immunodeficiency Syndrome Tobacco Use Disorder
chr5	58301612	59819647	134	PDE4D	Obesity, Asthma, Apoplexy, Stroke
chr9	89302478	89513369	121	DAPK1	colorectal cancer Alzheimer's disease
chr13	23893069	23984948	118	PARP4	Bladder cancer Chronic obstructive pulmonary
chr2	234191030	234346690	117	UGT1A	Colorectal cancer Chronic renal failure
chr13	41520889	41728716	114	DGKH	Tobacco Use Disorder Bipolar Disorder
chr11	119713156	119865855	103	ARHGEF12	HIV Infections, Longevity
chr6	101953675	102623474	102	GRIK2	Autism Tobacco Use Disorder
chr8	31616810	32741615	98	NRG1	Marijuana Abuse, Schizophrenia
chr7	39012933	39470915	98	POU6F2	autism, Alcoholism
chr16	17103682	17472239	95	XYLT1	Hypertension, Aortic Aneurysm
chr2	215301522	215382611	92	BARD1	Esophageal adenocarcinoma Breast cancer
chr1	206262211	206484288	92	PLXNA2	Anxiety Schizophrenia
chr16	23754823	24139431	90	PRKCB1	Type 2 diabetes Epilepsies
chr21	35081968	36278917	89	RUNX1	Leukemia Prostate cancer
chr16	52295376	52705882	88	FTO	Posttransplantation diabetes mellitus Obesity
chr1	229829184	230243641	88	DISC1	Marijuana Abuse Alzheimer's disease
chr11	107598769	107745036	87	ATM	Chronic renal failure Breast cancer
chr5	168025858	168660554	86	SLIT3	Tobacco Use Disorder Cervical Neoplasm
chr16	67328696	67426945	84	CDH1	Colocteral Cancer Retinal Diseases
chr12	116135362	116283965	79	NOS1	Marijuana Abuse, Malaria
chr5	42459783	42757683	78	GHR	Dwarfism, Colorectal cancer
chr2	40192790	40593079	78	SLC8A1	Hyperparathyroidism Tobacco Use Disorder
chr21	45649525	45758062	76	COL18A1	Head and Neck Neoplasms Hepatopulmonary Syndrome
chr16	9762923	10184112	76	GRIN2A	Narcolepsy Tobacco Use Disorder
chr7	107575318	107884062	75	NRCAM	Tobacco Use Disorder Autism
chr19	53310515	53365372	73	LIG1	Lymphoma, Non-Hodgkin Leukemia
chr20	17154752	17413222	71	PCSK2	Obesity Tobacco Use Disorder
chr7	134114711	134306012	70	CALD1	nephropathy in other diseases Diabetes, type 1

Table 1. Examples of GAD genes that contain large numbers of additional phased SNPs

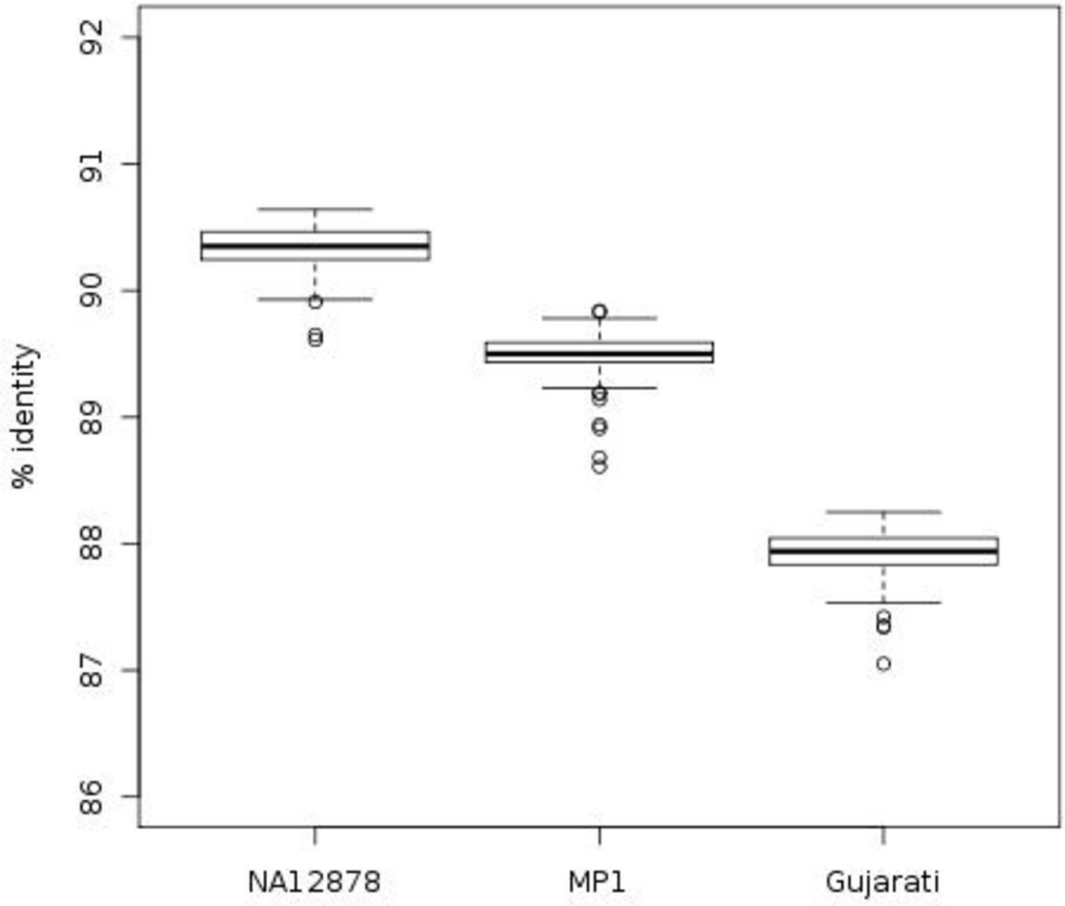

Figure 1. Distribution of concordance between haplotypes assembled with fosmids sequencing and 83 haplotypes assembled by trio phasing in the CEU HapMap population. NA12878 shares on average more heterozygous positions (145,187) than MP1 (134,947) and the Gujarati individual (111,818). Also, the percentage of concordance in the phasing of consecutive variants is on average larger for NA12878 (90.5%) and for MP1 (89.48%) than for the Gujarati individual (87.91%)

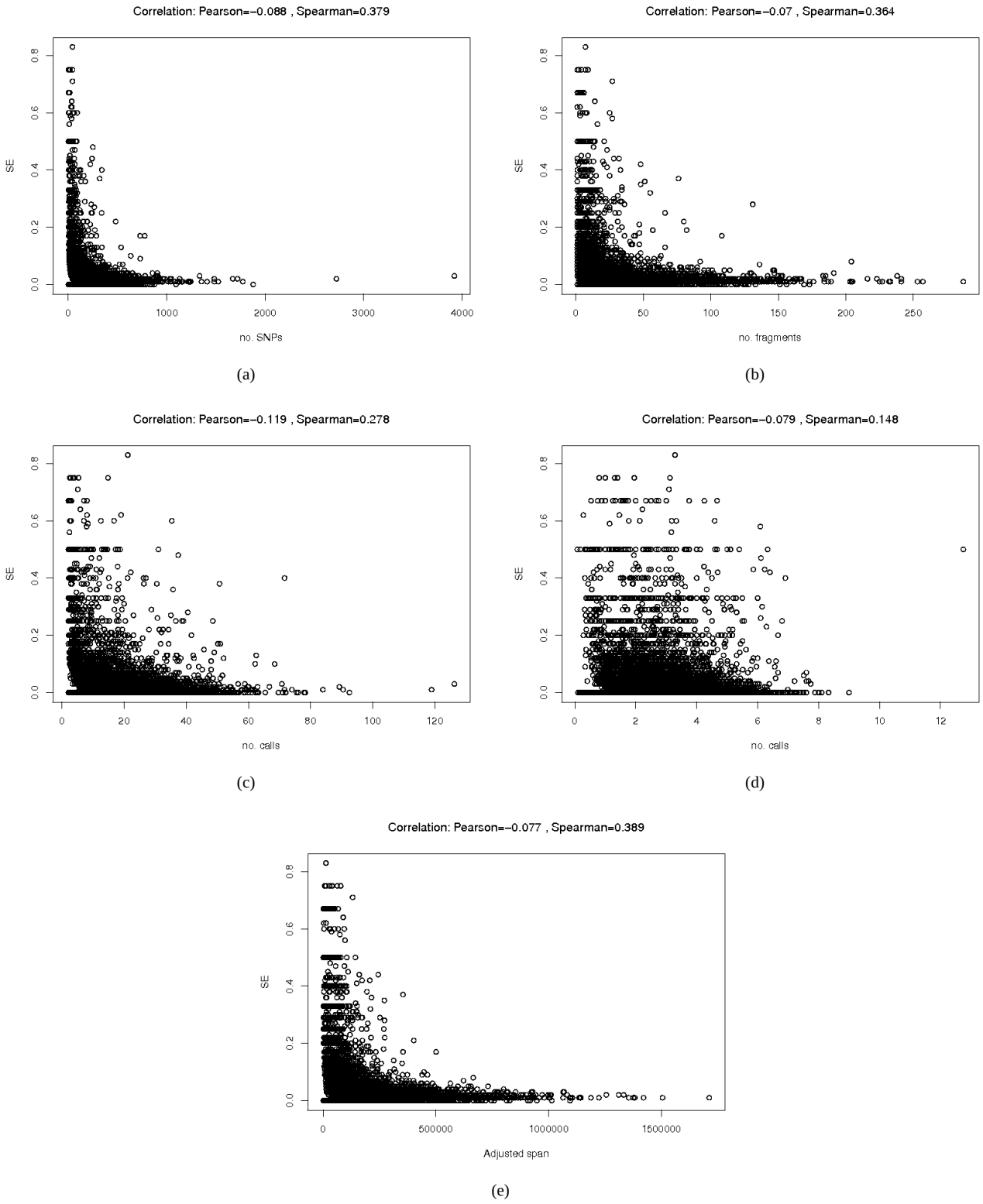

Figure 2. Distribution of switch error rate for blocks with different number of SNPs (a), number of fragments (b), average fragment length (c), average coverage (d), adjusted span (e)