Kidd Laboratory (January 2009) list of candidate SNPs for individual identification (IISNPs)

92 SNPs with average heterozygosity ≥ 0.4 and Fst(44pops) <0.06

Including a suggested set of 45 IISNP markers that are also "unlinked"

The SNPs are sorted by the Fst value based on a total of 44 population samples (more than 2,200 individuals typed). Four new population samples have been tested on all the SNPs screened since the "Provisional List of Candidates, Summer 2007"†. The 4 new samples include the Sandawe (East Africa), Hungarians (Europe), Keralites (South Central Asia), and Laotians (Southeast Asia). One of the original 40 best SNPs was dropped from the list after the expanded population testing due to an Fst>0.060; for convenience that marker is identified at the bottom of the table. Note that under the column labeled *Fst(44p) ranks* the aqua-blue highlighting of ranks indicate markers that were in the original published list of 40 best SNPs (Pakstis et al., 2007). Markers studied by the SNPforID consortium (Sanchez et al., 2006) have a single asterisk tag after the Fst(44p) rank. Publications describing the identification of all but the most recently screened SNPs can be found in the appended citation list.

In column 1 of the table ("unlinked" IISNPs), the green-highlighted check marks (√) indicate 45 SNPs among the 92 candidate SNPs that appear to be the most useful for individual identification at this time; 33 of these 45 proposed IISNPs are more than 95 cM apart while the other 12 SNPs in the list of "unlinked" SNPs range from 41 to 94 cM apart. The 45 proposed IISNPs are spread across the 22 autosomes. The set of "unlinked" SNPs might still need adjustment depending on the typing procedures developed for the implementation of this recommended panel. For example, it may not be possible to include all 45 SNPs due to multiplexing problems. Substitute SNPs may be needed and the additional SNPs in the list below offer some alternate candidates on various chromosomes. All 92 SNPs meet the population genetics criteria (Fst<0.06 and average heterozygosity >0.4); however, genetic map distances for substitute SNPs on the IISNP list need to be considered carefully to avoid markers that are too closely linked and that thus may have a degree of linkage disequilibrium that renders the substitute marker too correlated with existing nearby IISNPs. In such a case the substitute SNP would not add a full marker's worth of independent information to the overall IISNP panel.

The table column labeled *Avg cM position* is a simple average of the centi-Morgan value of the polymorphism on the DeCode, Genethon, and Marshfield genetic maps (which were obtained from the NCBI Map Viewer). The reader is reminded that each of these extensive maps does not necessarily have the same starting point on each chromosome and that the density of markers will vary in different chromosome regions. The starting or zero positions are near the pter end of each chromosome.

Except for some of the most recently screened markers, the information here was included in figures and tables presented in posters at various scientific meetings. (See footnote †.) PDF files of the poster presentations and of the earlier preliminary candidate list as of summer 2007 can be found at the following "contents" web page under the Kidd Lab Library header (http://info.med.yale.edu/genetics/kkidd/contents.html).

Allele frequency tables for all 92 best candidate SNPs have been deposited into ALFRED, the Allele Frequency Database. ALFRED is freely accessible on the web at http://ALFRED.med.yale.edu. Allele frequency tables for several hundred SNPs that were screened for this low Fst—high heterozygosity project are in the process of being entered into ALFRED; many of these SNPs did not pass beyond the early screening stage in which they were typed for 7 population samples representing the major continental regions of the world.

√ marks 45 “unlinked” IISNPs; # indicates one of 40 best SNPs (Pakstis et al., 2007); * next to Fst rank tags SNPforID marker

unlinked IISNPs	Fst (44p) rank	TaqMan Catalog ID	dbSNP rs#	ALFRED UID	Avg.Het. (44p)	Fst (44p)	Chr	Chr arm	Nucleotide Position Map Build 36.2	Avg cM posi-tion
√	1	C___2450075_10	rs10488710	SI001899B	0.442	0.0217	11	q	114,712,386	111.6
√	2	C__16156638_10	rs2920816	SI015053O	0.459	0.0232	12	q	39,149,319	57.9
√	3	C__29220288_10	rs6955448	SI015041L	0.421	0.0298	7	p	4,276,891	7.6
√	4	C___1619935_1_	rs1058083	SI001402H	0.464	0.0300	13	q	98,836,234	84.6
√	5	C____824925_10	rs221956	SI015402M	0.462	0.0310	21	q	42,480,066	54.6
√	6	C___2556113_10	rs13182883	SI001390N	0.472	0.0314	5	q	136,661,237	140.6
√	7	C___8263011_10	rs279844	SI001391O	0.484	0.0316	4	p	46,024,412	61.8
√	8	C__11245682_10	rs6811238	SI001910L	0.484	0.0319	4	q	169,900,190	166.9
√	9	C___9603287_10	rs430046	SI015042M	0.441	0.0321	16	q	76,574,552	94.1
√	10	C____788229_10	rs576261	SI015043N	0.472	0.0352	19	q	44,251,647	63.6
	11	C__16071557_10	rs2833736	SI015401L	0.460	0.0356	21	q	32,504,593	32.2
√	12	C___2049946_10	rs10092491	SI001900K	0.459	0.0364	8	p	28,466,991	52.5
√	13	C___1006721_1_	rs560681	SI001392P	0.434	0.0364	1	q	159,053,294	167.3
	14	C___1056251_10	rs590162	SI015390S	0.482	0.0366	11	q	121,701,199	124.6
√	15	C___9084395_10	rs2342747	SI015395X	0.423	0.0367	16	p	5,808,701	10.1
√	16	C__26449463_10	rs4364205	SI015054P	0.458	0.0372	3	p	32,392,648	56.3
√	17	C___2997607_10	rs445251	SI001912N	0.464	0.0386	20	p	15,072,933	36.8
√	18	C__29060279_10	rs7041158	SI015389A	0.439	0.0389	9	p	27,975,938	51.3
	19	C___1797119_10	rs9546538	SI003897B	0.429	0.0395	13	q	83,354,736	69.6
√	20	C___1304451_10	rs1294331	SI015382T	0.457	0.0396	1	q	231,515,036	247.4
√	21	C___1454681_10	rs159606	SI015134O	0.442	0.0396	5	p	17,427,898	23.8
√	22	C___3254784_10	rs740598	SI001393Q	0.462	0.0406	10	q	118,496,889	139.1
	23	C___3031045_1_	rs464663	SI015400K	0.462	0.0410	21	q	26,945,241	25.7
√	24	C__11673733_10	rs1821380	SI001913O	0.465	0.0413	15	q	37,100,694	38.2
√	25	C___1817429_10	rs1336071	SI001915Q	0.472	0.0418	6	q	94,593,976	102.3
	26	C___2572254_10	rs1019029	SI001916R	0.474	0.0419	7	p	13,860,801	23.0
√	27	C___1371205_10	rs9951171	SI001395S	0.475	0.0420	18	p	9,739,879	31.4
√	28	C___7968314_10	rs8078417	SI015122L	0.402	0.0426	17	q	78,055,224	130.0
	29	C___2140539_10	rs1358856	SI001427O	0.474	0.0430	6	q	123,936,677	121.3
√	30	C__25749280_10	rs6444724	SI001903N	0.469	0.0435	3	q	194,690,074	217.4
√	31	C___9371416_10	rs13218440	SI001397U	0.458	0.0436	6	p	12,167,940	24.6
	32	C__15957782_10	rs2270529	SI015388Z	0.421	0.0443	9	p	14,737,133	28.9
√	33	C___1452175_	rs1498553	SI015123M	0.477	0.0446	11	p	5,665,604	11.4
√	34	C____342791_10	rs7520386	SI001394R	0.477	0.0447	1	p	14,027,989	29.7
√	35	C___2508482_10	rs1523537	SI001914P	0.472	0.0447	20	q	50,729,569	79.4
√	36	C___3285337_	rs1736442	SI015124N	0.438	0.0450	18	q	53,376,775	79.4
	37	C___1152009_10	rs1478829	SI001917S	0.474	0.0459	6	q	120,602,393	119.8

√	38	C___2822618_10	rs3780962	SI001904O	0.476	0.0462	10	p	17,233,352	42.7
	39	C____105475_10	rs7229946	SI001901L	0.464	0.0466	18	q	20,992,999	49.8
	40	C__30281961_10	rs9866013	SI015044O	0.419	0.0468	3	p	59,463,380	77.4
	41	C___3206279_1_	rs2567608	SI001902M	0.473	0.0469	20	p	22,965,082	49.8
√	42	C___1541359_10	rs2399332	SI015385W	0.435	0.0472	3	q	111,783,816	124.5
√	43	C__11887110_1_	rs987640	SI001918T	0.476	0.0476	22	q	31,889,508	34.9
	44	C____376875_10	rs4847034	SI015135P	0.445	0.0476	1	p	105,519,154	134.1
	45	C__11522503_1_	rs2073383	SI001911M	0.456	0.0479	22	q	22,132,171	15.8
	46	C__26227271_10	rs3744163	SI015125O	0.430	0.0480	17	q	78,333,148	130.0
	47	C___1605841_10	rs10500617	SI003936V	0.404	0.0481	11	p	5,055,969	9.0
√	48	C___9530932_10	rs993934	SI015136Q	0.450	0.0482	2	q	123,825,683	134.2
	49	C___7969752_	rs2291395	SI015126P	0.473	0.0486	17	q	78,119,428	130.0
√	50	C___2715242_10	rs10773760	SI015392U	0.444	0.0487	12	q	129,327,649	165.1
	51	C___1274218_	rs12480506	SI001169R	0.403	0.0492	20	p	16,189,416	39.1
	52	C__11258596_	rs4789798	SI015127Q	0.472	0.0494	17	q	78,124,932	130.0
√	53	C____187613_10	rs4530059	SI015393V	0.406	0.0495	14	q	103,840,194	126.5
	54	E_rs8070085_10	rs8070085	SI014994B	0.437	0.0498	17	q	38,595,510	66.4
√	55	C___1276208_10	rs12997453	SI001396T	0.440	0.0503	2	q	182,121,504	188.1
√	56	C__27999762_10	rs4606077	SI015387Y	0.421	0.0503	8	q	144,727,897	164.2
	57	C______19853_	rs689512	SI001329P	0.423	0.0507	17	q	78,308,991	130.0
√	58	C___2515223_10	rs214955	SI001403I	0.474	0.0511	6	q	152,739,399	155.7
	59	C___1256256_1_	rs2272998	SI001398V	0.467	0.0511	6	q	148,803,149	148.6
	60	C___2539254_	rs5746846	SI003887A	0.464	0.0515	22	q	18,300,646	9.0
	61	C__26372385_10	rs4288409	SI015386X	0.415	0.0515	8	q	136,908,411	152.0
√	62	C___2184724_	rs2269355	SI015128R	0.473	0.0521	12	p	6,816,175	17.0
	63	C___1570295_10	rs1027895	SI000905O	0.433	0.0524	17	q	43,865,696	69.4
√	64	C___3004178_10	rs321198	SI001906Q	0.459	0.0530	7	q	136,680,378	143.5
	65	C__11631183_	rs2175957	SI015129S	0.437	0.0530	17	q	38,540,348	66.3
	66	C___3080506_1_	rs2292972	SI001330H	0.422	0.0530	17	q	78,359,077	130.0
	67 *	C___7698393_	rs901398	SI003975Y	0.441	0.0531	11	p	11,052,797	18.2
	68	C___2539253_	rs9606186	SI001586U	0.437	0.0531	22	q	18,300,359	9.0
√	69	C__3153696a_10	rs338882	SI001401G	0.469	0.0532	5	q	178,623,331	195.8
√	70	C___2002375_10	rs10776839	SI015046Q	0.463	0.0533	9	q	136,557,129	152.6
	71	C___2714437_	rs521861	SI001163L	0.473	0.0534	18	q	45,625,012	70.7
√	72	C___2073009_10	rs1109037	SI001909T	0.470	0.0534	2	p	10,003,173	21.5
	73	C__29487208_10	rs4796362	SI015397Z	0.471	0.0536	17	p	6,752,253	14.2
	74	C___3032822_1_	rs315791	SI001404J	0.472	0.0539	5	q	169,668,498	176.3
	75 *	C___7539584_	rs891700	SI003976Z	0.471	0.0541	1	q	237,948,549	261.3
	76	C___7477802_	rs1004357	SI015131L	0.411	0.0541	17	q	39,047,052	67.1
	77	E_rs7205345_10	rs7205345	SI001905P	0.469	0.0544	16	p	7,460,255	14.2
	78	C__1636106a_10	rs6591147	SI001409O	0.451	0.0545	11	q	105,418,194	106.3
	79	C____411273_10	rs2503107	SI001426N	0.458	0.0548	6	q	127,505,069	125.9
	80	C___7538108_10	rs1410059	SI001399W	0.470	0.0551	10	q	97,162,585	117.6
	81	C__11907549_1_	rs1872575	SI003924S	0.472	0.0552	3	q	115,287,669	128.2

	82	C___7428940_10	rs1554472	SI001919U	0.472	0.0552	4	q	157,709,356	155.7
	83 *	C__11989432_10	rs2046361	SI003977A	0.462	0.0559	4	p	10,578,157	23.1
√	84	C___7945874_10	rs9905977	SI015045P	0.419	0.0561	17	p	2,866,143	7.9
	85	C___1995608_10	rs7704770	SI001908S	0.449	0.0567	5	q	159,420,531	163.0
	86	C___1880371_10	rs13134862	SI001400F	0.453	0.0571	4	q	76,644,920	84.2
	87	C____282853_10	rs2811231	SI015137R	0.458	0.0579	6	p	55,263,663	78.9
	88	C___7459903_10	rs985492	SI001413J	0.469	0.0580	18	q	27,565,032	58.6
	89	C___1605842_	rs10768550	SI003937W	0.408	0.0580	11	p	5,055,290	9.0
	90 *	C___9630073_	rs1490413	SI003978B	0.469	0.0583	1	p	4,267,183	8.3
	91	C__11338582_	rs2255301	SI001069Q	0.463	0.0587	12	p	6,779,703	16.9
√	92	C____611046_10	rs722290	SI003542O	0.468	0.0596	14	q	52,286,473	47.6

SNP below dropped from IISNP list when Fst exceeded 0.06 after expanding to 44 population samples

XXX	C___2223883_10	rs447818	SI001907R	0.469	0.0622	6	q	145,910,689	145.1

Note:

Only chance level linkage disequilibrium (LD) values are observed for all unique pairings of 86 of the 92 IISNPs (median LD =0.011) in each of 44 population samples. However, six of the 92 SNPs show strong LD in most of the 44 populations for a small subset (7) of the unique pairings due to close linkage; these 6 SNPs can therefore only be alternative candidates for inclusion in an applied IISNP panel of 86 SNPs independent at the population level. These six SNPs showing some LD are those in the above table with Fst ranks numbered: 52, 57, 65, 66, 68, and 89.

Citations

Pakstis et al. 2007. *Human Genetics* 121:304-317. A PDF file of this paper (publication #461) can be downloaded at: http://info.med.yale.edu/genetics/kkidd/pubs.html. (See also publications 467 and 468.)

Pakstis et al. 2009. *Human Genetics*, DOI 10.1007/s00439-009-0771-1. Published online at journal website Nov 25, 2009.

Sanchez et al. 2006. *Electrophoresis* 27:1713-1724.

† Scientific meetings where much of this information was presented:

Figures 1 and 2 of Poster presentation July 24, 2007 for the annual meeting of grantees of theU.S. National Institue of Justice, Washington, D.C.
Title: An expanded, nearly universal, panel of SNPs for individual identification.
Authors: Andrew J. Pakstis, William C. Speed, Judith R. Kidd, Kenneth K. Kidd

Affiliation: Dept of Genetics, Yale University School of Medicine, New Haven, CT

Figure 1 of Poster presentation August 22-25, 2007 for the meeting of the International Society of Forensic Geneticists (ISFG) in Copenhagen, Denmark.
Title: SNPs for individual identification
Authors: Andrew J. Pakstis, William C. Speed, Judith R. Kidd, Kenneth K. Kidd
Affiliation: Dept of Genetics, Yale University School of Medicine, New Haven, CT

Poster presentation July 21-23, 2008 for the annual meeting of grantees of the U.S. National Institute of Justice, Arlington, Virginia .
Title: Better panels of SNPs for ancestry inference and individual identification
Authors: Andrew J. Pakstis, William C. Speed, Judith R. Kidd, Kenneth K. Kidd
Affiliation: Dept of Genetics, Yale University School of Medicine, New Haven, CT

Poster presentation June 15-17, 2009 for the annual meeting of grantees of the U.S. National Institute of Justice, Arlington, Virginia.
Title: SNP panels for individual identification and for ancestry inference
Authors: Kenneth K. Kidd, Judith R. Kidd, William C. Speed, Andrew J. Pakstis
Affiliation: Dept of Genetics, Yale University School of Medicine, New Haven, CT

Acknowledgments

In addition to direct funding from grants (see below) we thank Applied Biosystems for providing some of the TaqMan assays used for screening SNPs on our population samples. We also thank the many hundreds of individuals who volunteered to provide samples for genetic marker testing.

Financial Support

This work was supported in part by multiple funding sources. The assembly of some of the population resources was funded in part by NIH grants (GM057672 and AA009379). The forensic aspects of the project were funded in part by the U.S. DOJ NIJ; specifically, this work was supported in part by awards 2004-DN-BX-K025 and 2007-DN-BX-K197 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication/program/exhibition are those of the authors and do not necessarily reflect those of the Department of Justice.