Large-Scale Zygosity Testing Using Single Nucleotide Polymorphisms

Ulf Hannelius,[1] Loreana Gherman,[3] Ville-Veikko Mäkelä,[2] Astrid Lindstedt,[2] Marco Zucchelli,[1] Camilla Lagerberg,[3] Gunnel Tybring,[3] Juha Kere,[1,2,4] and Cecilia M Lindgren[5]

[1] *Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden*
[2] *Clinical Research Centre, Karolinska University Hospital, Stockholm, Sweden*
[3] *Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden*
[4] *Department of Medical Genetics, University of Helsinki, Helsinki, Finland*
[5] *Wellcome Trust Centre for Human Genetics, Oxford University, Oxford, United Kingdom*

A requirement for performing robust genetic and statistical analyses on twins is correctly assigned zygosities. In order to increase the power to detect small risk factors of disease, zygosity testing should also be amenable for high throughput screening. In this study we validate and implement the use of a panel of 50 single nucleotide polymorphisms (SNPs) for reliable high throughput zygosity testing and compare it to a panel of 16 short tandem repeats (STRs). We genotyped both genomic (gDNA) and whole genome amplified DNA (WGA DNA), ending up with 47 SNP and 11 STR markers fulfilling our quality criteria. Out of 99 studied twin pairs, 2 were assigned a different zygosity using SNP and STR data as compared to self reported zygosity in a questionnaire. We also performed a sensitivity analysis based on simulated data where we evaluated the effects of genotyping error, shifts in allele frequencies and missing data on the qualitative zygosity assignments. The frequency of false positives was less than 0.01 when assuming a 1% genotyping error, a decrease of 10% of the observed minor allele frequency compared to the actual values and up to 10 missing markers. The SNP markers were also successfully genotyped on both gDNA and WGA DNA from whole blood, saliva and filter paper. In conclusion, we validate a robust panel of 47 highly multiplexed SNPs that provide reliable and high quality data on a range of different DNA templates.

By making use of the genetic variation present in the human genome it is possible to obtain unique genetic fingerprints that can discriminate between individuals (Jeffreys et al., 1985). This also offers the most robust method for estimating the zygosity of a twin pair (Jackson et al., 2001; Ooki et al., 2004; Reed et al., 2005). In order to obtain a reliable genetic fingerprint it is important to consider issues relating to the quality and quantity of the DNA samples available, as well as the choice of genetic markers. First, the most common way of obtaining a unique genetic fingerprint is to genotype a set of highly polymorphic short tandem repeats (STRs). However, because of their relatively large amplicon sizes compared to both single nucleotide polymorphisms (SNPs) and mini-STRs, they are prone to amplification failure and allelic imbalance due to degraded DNA (Dixon et al., 2005; Petkovski et al., 2005; Utsuno & Minaguchi 2004). Second, the use of whole genome amplification (WGA), where the genome is amplified in order to increase the life-span and usage of individual DNA samples, has raised concerns relating to biased amplification of one allele over the other depending on the template amount and WGA method (Dean et al., 2002; Lovmar et al., 2003). This problem is also more pronounced when using STRs over SNPs for subsequent genotyping (Bergen et al., 2005). Third, zygosity testing should optimally be amenable for high-throughput and cost-efficient screening, in order to avoid low-powered studies and support the ever expanding twin registries and collaboration efforts (Hirschhorn et al., 2002; Ioannidis et al., 2001; Ioannidis et al., 2003; Lohmueller et al., 2003; Peltonen 2003).

In this study we set out to evaluate and improve an already published set of 41 SNP markers (Petkovski et al., 2005) ,and to validate the modified protocol for zygosity testing on genomic (gDNA), as well as WGA DNA ,using the REPLI-g kit (Dean et al., 2002; Hosono et al., 2003) on DNA extracted from whole blood and saliva. We show the utility of a highly multiplexed SNP panel for high-throughput zygosity testing and general fingerprinting, and compare it to a panel of STRs. We also consider the use of different DNA templates and WGA in the context of genetic fingerprinting and further evaluate the robustness of the SNP panel through simulation studies.

Received 12 April, 2007; accepted 23 May, 2007.

Address for correspondence: Ulf Hannelius, Department of Biosciences and Nutrition, Karolinska Institutet, Hälsovägen 7–9, 14157 Huddinge, Sweden E-mail: ulf.hannelius@biosci.ki.se

Table 1

SNP Marker Information and Allele Frequencies Including 95% Confidence Intervals for the CEU HapMap Population and Internally Genotyped Samples

SNP id[a]	ch[r]	Locus[a]	Pos[a]	allele[s]	CEU[b]	Internal[c]
AMELXY	X/Y	AMEL	NA	NA	NA	NA
rs1020636	2	LTBP1	33097912	T/C	0.62(0.49–0.74)	0.63(0.58–0.68)
rs1111366	1	LOC339505	20463069	C/A	0.56(0.43–0.69)	0.60(0.55–0.66)
rs11249784	5	LOC441119	177316388	A/G	0.65(0.53–0.77)	0.69(0.64–0.74)
rs11706962	3	—	196061467	G/A	0.62(0.50–0.75)	0.58(0.53–0.64)
rs1361861	X	IL1RAPL2	103855883	G/A	0.52(0.39–0.64)	0.55(0.48–0.63)
rs1403294	8	—	104414954	T/C	0.67(0.56–0.79)	0.73(0.67–0.78)
rs1479530	3	CNTN6	1375419	C/G	0.62(0.50–0.75)	0.64(0.58–0.69)
rs1500098	12	IQSEC3	59354	C/G	0.37(0.24–0.49)	0.46(0.41–0.52)
rs1620329	11	OPCML	132082653	C/T	0.39(0.27–0.52)	0.43(0.37–0.49)
rs16282	X	LOC203427	118362512	C/T	0.33(0.21–0.45)	NA
rs1674139	19	AP2A1	54963314	T/C	0.38(0.26–0.51)	0.43(0.37–0.49)
rs17379	X	H6ST2	131802447	A/G	0.57(0.45–0.70)	0.51(0.43–0.59)
rs17407	X	—	140808205	A/G	0.62(0.50–0.75)	0.67(0.60–0.75)
rs1860665	9	—	119501589	T/G	0.46(0.33–0.58)	0.51(0.45–0.57)
rs1894697	1	F5	166258259	G/C	0.50(0.37–0.63)	0.49(0.44–0.55)
rs1924609	13	ATP7B	51410519	C/T	0.49(0.37–0.62)	0.49(0.44–0.55)
rs1936827	X	—	125063418	G/A	0.63(0.51–0.76)	NA
rs222	2	INPP4A	98606390	T/C	0.76(0.65–0.87)	0.79(0.74–0.83)
rs228043	21	SLC37A1	42823507	A/G	0.51(0.38–0.64)	0.53(0.47–0.59)
rs2282739	1	HSD11B1	206271396	C/T	0.74(0.62–0.85)	0.70(0.65–0.76)
rs2289105	15	CYP19A1	49294800	T/C	0.41(0.28–0.53)	0.52(0.46–0.58)
rs230	4	—	173271203	A/G	0.39(0.27–0.52)	0.47(0.41–0.52)
rs2303025	5	ANXA6	150483868	T/C	0.56(0.43–0.69)	0.59(0.54–0.65)
rs234	7	—	105155086	T/C	0.56(0.43–0.68)	0.53(0.47–0.58)
rs240	9	—	109558585	C/G	0.55(0.42–0.68)	0.63(0.57–0.68)
rs276922	18	DSC3	26858793	A/C	0.57(0.44–0.69)	0.52(0.46–0.57)
rs326414	3	—	7878040	A/G	0.55(0.42–0.68)	0.52(0.47–0.58)
rs3784740	15	ST8SIA2	90761479	T/C	0.55(0.42–0.68)	0.55(0.50–0.61)
rs4240868	1	—	149794672	G/A	0.52(0.39–0.64)	0.46(0.40–0.51)
rs4306954	4	—	105660759	T/C	0.53(0.40–0.65)	0.49(0.43–0.55)
rs4358717	7	—	140396256	T/C	0.57(0.44–0.70)	0.56(0.51–0.62)
rs4763188	12	PPM1H	61333020	T/A	0.59(0.47–0.72)	0.65(0.60–0.70)
rs544021	11	ACTN3/ZDHHC24	66072237	C/G	0.57(0.44–0.69)	0.55(0.49–0.61)
rs6115	14	SERPINA5	94123643	G/A	0.38(0.26–0.51)	0.35(0.29–0.40)
rs6771379	3	—	5590030	C/T	0.32(0.21–0.44)	0.36(0.30–0.41)
rs710891	16	TRAP1	3650098	C/T	0.65(0.53–0.77)	0.60(0.54–0.66)
rs724784	5	MAN2A1	109081234	A/C	0.66(0.54–0.78)	0.61(0.56–0.67)
rs754	20	—	58102499	A/G	0.92(0.85–0.99)	0.87(0.84–0.91)
rs7747651	6	EFHC1	52433769	C/A	0.63(0.51–0.76)	NA
rs7994365	13	AKAP11	41765954	A/G	0.42(0.30–0.55)	0.45(0.39–0.50)
rs811	12	—	89278405	G/A	0.35(0.23–0.47)	0.35(0.30–0.41)
rs820129	17	SAP30BP	71179702	A/G	0.67(0.55–0.79)	0.65(0.60–0.70)
rs874746	8	KCNK9	140722490	C/A	0.45(0.32–0.58)	0.41(0.36–0.47)
rs882937	11	PANX1	93548520	A/G	0.44(0.32–0.57)	0.44(0.39–0.50)
rs889012	5	TRPC7	135582615	C/A	0.48(0.36–0.61)	0.55(0.50–0.61)
rs910170	6	GLP1R	39140393	T/C	0.44(0.32–0.57)	0.45(0.40–0.51)
rs9663989	10	TLL2	98225107	A/C	0.39(0.27–0.52)	0.44(0.39–0.50)
rs9788905	16	—	65000382	T/C	0.61(0.48–0.73)	0.63(0.58–0.68)
rs997556	7	—	51738954	T/C	0.69(0.57–0.81)	0.59(0.54–0.65)

Note: [a]The reference SNP id:s, chromosomal positions and gene symbols are presented according to NCBI build 35.

[b]30 trios from Utah Residents with Northern and Western European Ancestry (CEU). The allele frequencies represent non-redundant HapMap data released on the July 20, 2006.

[c]15 CEU trios, 24 unrelated individuals from the Coriell Institute and 99 randomly chosen individuals from each genotyped twin pair.

For X-chromosomal markers the allele frequencies are calculated using only female individuals. All frequencies are presented in relation to the HapMap reference allele.

The reference allele is presented as the first allele in the alleles-column. SNPs rs16282, rs1936827 and rs7747651 failed our quality criteria.

https://doi.org/10.1375/twin.10.4.604 Published online by Cambridge University Press

Materials and Methods

Samples

The study was approved by the regional ethical review board at Karolinska Institutet. gDNA from 14 trios and 24 unrelated individuals from Utah Residents with Northern and Western European Ancestry was obtained from the Coriell Institute (Appendix A, Supplementary Table 1,). Additionally, gDNA from 198 twin samples was obtained from the Swedish Twin Registry (Lichtenstein et al., 2002) and gDNA from 11 randomly chosen unrelated donors was collected in the form of whole blood, saliva and as blood spots on filter paper. For details concerning DNA extraction and quantification see Appendix A, Supplementary Methods.

Whole Genome Amplification

10 ng DNA was used for all WGA reactions. The DNA from the 14 trios and 24 unrelated Coriell samples were amplified using the REPLI-g Mini kit, while the DNA from the 11 unrelated donors were amplified using the REPLI-g Midi kit. Both reactions were performed according to manufacturer's protocol. Amplification of the DNA extracted from filter paper was performed according to the February 2005 supplementary protocol for filter paper samples.

Genotyping — Design of SNP Panel

The 41 SNPs as described by Petkovski et al. (Petkovski et al., 2005) were used as a starting point for the design of the SNP panel. All non-HapMap SNPs (Phase I) were excluded, and SNPs genotyped by HapMap with minor allele frequencies greater than 20%, and genotypes distributed according to HWE in trios from Utah Residents with Northern and Western European Ancestry (CEU), were added, until a panel of 50 markers with no inter-marker LD, including the sex specific AMELXY marker, was obtained (Table 1). SNPHunter (Wang et al., 2005) and RepeatMasker version open-3.1.5 (http://www.repeatmasker.org/) were used for detecting nearby SNPs and repeats in the flanking sequences, and Sequenom's Assay Design 3.4 software was used for designing multiplexes for the iPLEX chemistry.

Genotyping — SNP Method

The PCR and subsequent downstream processing was performed according to the manufacturer's standard protocol (Sequenom), and the genotypes were analyzed by an Autoflex MassARRAY mass spectrometer (Bruker Daltonics, Billerica, MA, USA). The data were analyzed independently by two persons using the SpectroTyper software (Sequenom Inc.). For details regarding genotyping and primer sequences see Appendix A, Supplementary Methods.

Genotyping — Design of STR Panel

The primer sequences in the initial multiplexing schemes were from the commercially available PowerPlex® 16 (ProMega) STR panel, as well as from primer sequences reported in UNISTS for the D19S433 (UniSTS:33588) and D2S1338 (UniSTS:30509) markers. These were used for designing a panel of 16 STR markers (Table 2). After initial validation (see below) the final panel was reduced to 15 STR markers assayed in three multiplexes.

PCR Protocol — Microsatellite Genotyping

The PCR was performed in a 384 well format and a total volume of 5 µl using Applied Biosystems thermocyclers (See Appendix A, Supplementary Methods for details). The microsatellites were run and analyzed on a MegaBace1000 capillary sequencer (GE Healthcare). The data were analyzed independently by two persons using the Genetic Profiler version 2.2 (GE Healthcare).

Table 2

STR Marker Information and Allele Frequencies Based on Internally Genotyped Samples

Locus	Chr	Locus	1	2	3	4	5	6	7	8	9	10	11	12	13
Amelogenina	X/Y	*AMEL*													
CSF1PO	5	*CSF1PO*	.26	.31	.33	.04	.01	.00	.01	.01	.02				
D13S317	13	—	.12	.08	.09	.24	.32	.11	.04						
D16S539	16	—	.12	.08	.29	.31	.18	.01	.01						
D18S51	18	*UT574*	.19	.17	.14	.12	.10	.03	.02	.00	.00	.01	.00	.13	.09
D31358	3	—	.27	.14	.02	.00	.00	.11	.25	.21					
D5S818	5	—	.06	.36	.35	.17	.01	.01	.00	.03					
D7S820	7	—	.17	.23	.24	.16	.04	.00	.02	.14					
D8S1179	8	—	.02	.11	.09	.13	.29	.25	.07	.03	.02				
FGA	4	*FIBRA*	.04	.08	.13	.19	.18	.13	.15	.07	.03	.01	.01		
TH01	11	*TH01*	.17	.14	.13	.32	.01	.24							
TPOX	2	*TPOX*	.01	.51	.09	.04	.27	.07	.01						

Note: The allele frequencies are calculated only for the 11 markers that passed our quality criteria. The frequencies are calculated using genotype data stemming from whole genome amplified DNA from 14 CEU trios, 24 unrelated individuals from the Coriell Institute and 99 randomly chosen individuals from each genotyped twin pair. Markers D19S433, D21S11, D2S1338 and vWA are not included in the table since they failed our quality criteria.

https://doi.org/10.1375/twin.10.4.604 Published online by Cambridge University Press

Characterization of Markers and Analysis of Genotyping Data

An in-memory data visualization tool (Qlikview, Qliktech Sweden) was used to build an in-house application around nonredundant allele frequency data (Data release #21) from the HapMap project (2003) as well as gene and reference SNP information from NCBI build 35 (www.ncbi.nlm.nih.gov). The application was used for characterizing the initial 41 SNPs based on their ref-SNP ID as well as for identifying additional SNPs based on their allele frequencies and chromosomal locations. Calculations for genotyping success rates, concordance as well as data conversions were done in Qlikview version 7.5 (Qliktech) by incorporating genotype data from the Sequenom database and genotype reports from Genetic Profiler version 2.2. Initial checks for Mendelian inconsistencies as well as Hardy Weinberg and linkage disequilibrium calculations were done in Haploview version 3.2 (Barrett et al., 2005). Hardy Weinberg p values reported here were calculated using the exact test implemented in PowerMarker version 3.25 (Guo & Thompson 1992; Liu & Muse 2005). The unadjusted α-level for HWE calculations was 0.05, and in order to account for multiple testing we used Bonferroni adjusted α-levels. We failed markers that had success rates below 80% or that were out of HWE when considering adjusted α-levels. STR allele frequencies were calculated using PowerMarker version 3.2 and SNP allele frequencies, including 95% confidence intervals, were calculated in Qlikview.

Zygosity Testing

The likelihood of zygosity for the genotyped twin samples were calculated based on observed genotypes using ECLIPSE version 1.1, incorporating a prior genotyping error rate of 1% according to error model 0 (Sieberts et al., 2002).

Simulation Studies

Genotyping data from four sets of 10,000 MZ and 10,000 DZ twin pairs each were simulated, incorporating a genotyping error (ε) of 0% for simulation I and III, and 1% for simulations II and IV (Table 3). The genotyping error was introduced conditional on the true genotype as presented in Table 4. The minor allele frequencies used in simulations III and IV were decreased by 10%, while no allele frequency changes were introduced for simulations I and II (Table 3).

Missing data were simulated by randomly failing markers for each simulated twin pair. For details regarding the simulations see Appendix A, Supplementary Methods. The generated datasets were analyzed using ECLIPSE2 as described previously. Based on a likelihood ratio cut-off value of 1, we then calculated the frequency of false positives for DZ and MZ twins. A twin pair was considered a false positive if the likelihood ratio (MZ/DZ) was above 1 in the case of a DZ twin pair, and below 1 in the case of a MZ twin pair.

Table 3

Settings for the Simulation Studies Regarding Genotyping Error, Allele Frequency Shifts, Missing Markers and Prior Error Rates.

	I	II	III	IV
Genotyping error	0%	1%	0%	1%
Allele frequency shifts	0%	0%	–10%	–10%
Missing markers (out of 43)	0, 5 and 10	0, 5 and 10	0, 5 and 10	0, 5 and 10
Prior error	1%	1%	1%	1%

Table 4

The Probability of a Genotyping Error Given the True Genotype as Modeled in the Simulation Studies

	True genotype		
False genotype	AA	Aa	aa
AA	—	$\varepsilon/2$	0
Aa	ε	—	ε
aa	0	$\varepsilon/2$	—

Note: At least one of the alleles must be correct, resulting in a homozygous genotype (AA or aa) that has a probability of ε to get the wrong genotype (Aa). Similarly, a heterozygous genotype (Aa) has a probability of $\varepsilon/2$ to get either homozygous genotype (AA or aa).

Results

Validation of the Marker Panels

In order to evaluate the robustness of all the 50 SNP (Table 1) and 16 STR (Table 2) markers, we performed genotyping experiments on gDNA and WGA DNA, and calculated the genotyping success rate and concordance. At each round of genotyping, we excluded markers that did not meet our quality criteria (Appendix A, Supplementary Tables 6 and 7), resulting in a final panel of 47 SNP and 11 STR markers and the additional sex specific assays. Two SNPs were rejected in the first round of genotyping, based on deviations from HWE and badly separated genotype clusters based on the allele specific peak heights. One additional SNP was rejected in the second round for a low genotyping success rate. One STR was rejected in the initial validation for a low success rate (< 80%), while two STRs were rejected in the second genotyping round for deviations from HWE. One additional STR was rejected in the second round due to low genotyping success rate (< 80%). The marker specific success rates and genotype concordances, as well as HWE p values for the SNP and the STR markers, are presented in Tables 6 and 7.

In order to validate the SNP panel on a range of different DNA templates, we extracted DNA from 11 samples each from whole blood, saliva and blood on filter paper. The extracted DNA samples were then whole genome amplified using the REPLI-g kit, and both the gDNA and WGA samples were genotyped in

https://doi.org/10.1375/twin.10.4.604 Published online by Cambridge University Press

duplicates for the 47 SNP markers. The overall genotype concordance was 99% (*SD* 2%) for the gDNA samples and 100% (*SD* 1%) for the WGA DNA samples (Appendix A, Supplementary Table 4). The gDNA extracted from filter paper resulted in the lowest genotyping success rate (66%, *SD* 13%) while the rest of the success rates ranged between 89% and 99%.

Zygosity Testing

Next we set out to determine the zygosity of the genotyped twin samples. In order to keep the genotyping and the quality control unbiased, the zygosity of the pairs and the gender of the individuals were unknown to the persons performing the experiments. After decoding the sample codes, we calculated the allele sharing and the odds of the likelihood ratios of a pair being monozygotic versus dizygotic for all the pairs, and compared our results with the information provided by the Swedish Twin Registry. We incorporated a prior genotyping error rate of 1% into the ECLIPSE calculations in order to account for incorrect genotypes, as well as possible somatic mutations. One pair failed the genotyping for the SNP panel and two pairs for the STR panel. Two pairs (numbers 11 and 73) gave conflicting results regarding zygosity when comparing both the SNP and STR panels to the previously assigned zygosity. Here, one of the pairs (number 11) was DZ according to the questionnaire, and MZ according to the marker panels, whereas the other pair (number 73) was DZ according to the questionnaire, and MZ according to the marker panels. The gender specific assay in the SNP panel confirmed the sex of all twin pairs, while the corresponding assay in the STR panel failed to generate successful genotypes for nine of the individuals (Appendix A, Supplementary Table 5). Details regarding success rates and likelihood ratios of zygosity are presented in Appendix A, Supplementary Table 5.

We then performed four simulations of 10,000 MZ and 10,000 DZ pairs each to determine how sensitive the SNP panel is to genotyping error, shifts in allele frequencies and missing data. We again considered genotyping error rates of up to 1% and minor allele frequency shifts of up to –10% (Table 3). When considering a single likelihood cut-off of 1 and no missing data, the false positive rates for DZ twins were 0.03%, 0.03%, 0.32% and 0.19% for simulations I, II, III and IV respectively. There were no false positives for MZ twins for all simulations when all markers were included. The same datasets were reanalyzed after randomly dropping out 5 or 10 of the markers, yielding false positive rates of up to 0.79% for the DZ twins and 0.02% for the MZ twins (Table 5).

Table 5

The Per Cent False Positives of Simulated MZ and DZ Twin Pairs for a Given Number of Missing Markers.

		Missing markers		
Simulation		0	5	10
I	MZ	0	0	0
	DZ	0.03	0.12	0.3
II	MZ	0	0	0.01
	DZ	0.03	0.15	0.23
III	MZ	0	0	0
	DZ	0.32	0.57	0.79
IV	MZ	0	0.01	0.02
	DZ	0.19	0.3	0.71

Discussion

Genotyping a set of polymorphic genetic markers yields the most robust estimates of zygosity (Jackson et al., 2001; Ooki et al., 2004; Peeters et al., 1998; Reed et al., 2005). Other methods, like validated questionnaires, or data on chorionicity, are more imprecise, and may lead to an inflated false positive rate in association studies (Boomsma et al., 2002; Reed et al., 2005). In this study we set out to modify and validate a set of 41 previously published SNP markers (Petkovski et al., 2005), by improving and expanding the panel to 50 markers, and including a sex specific AMELXY assay (Table 1). Because zygosity testing and genetic fingerprinting are still often performed using STRs, we also compared the results of the SNP panel with a panel of 16 noncommerical STRs that are based on the Powerplex® 16 (Promega) and AmpFLSTR® Identifiler® (Applied Biosystems) panels (Table 2). Previous studies have demonstrated that STR genotyping can suffer from allele drop outs or total failure of amplification due to degraded DNA or WGA DNA (Barber & Foran 2006; Bergen et al., 2005; Dixon et al., 2005; Petkovski et al., 2005). Here we genotyped the two marker panels on both genomic and whole genome amplified DNA, and a total of 47 SNPs and 11 STRs passed our quality criteria (Tables 3 and 4). The fact that the STR calling requires more manual work probably explains partly the larger variation in both success rates and concordances compared to the SNP results (Appendix A, Supplementary Tables 2 and 3).

Applicability Using Different Templates

The use of optional sources for DNA, like saliva, blood on filter paper or WGA, has been shown to increase the response rate in epidemiological studies, and to facilitate the use of otherwise inaccessible biological repositories (Hannelius et al., 2005; Rylander-Rudqvist et al., 2006). There is consequently merit in applying similar strategies when expanding twin registries or replenishing exiting DNA samples, and panels of genetic markers that are used for zygosity testing should be compatible with these kinds of DNA templates. To explore this using our SNP panel, we genotyped 11 gDNA and WGA DNA samples extracted from whole blood, saliva and blood on filter paper, using our panel of 47

https://doi.org/10.1375/twin.10.4.604 Published online by Cambridge University Press

SNPs. The best genotyping success rates were obtained for the gDNA from whole blood and saliva, as well as for all three WGA DNA template sources. The gDNA filter paper template worked poorly, having an overall success rate of 66% and a *SD* of 13% (Appendix A, Supplementary Table 4). This is not surprising, because the method used for extracting DNA from filter paper results in at least partly fragmented DNA and low yield (Hannelius et al., 2005). The REPLI-g kit uses its own method for extracting DNA from filter paper, and it is therefore not possible to evaluate how well the WGA reaction improved on the quality of the gDNA extracted from filter paper DNA. Previous studies concerning different WGA methods have shown that kits like REPLI-g and GenomiPhi that are based on multiple strand displacement amplification (MDA) are superior when it comes to DNA of high molecular weight. Other methods like I-PEP-L and GenomePlex prevail when degraded DNA is used as template (Dean et al., 2002; Hannelius et al., 2005; Lovmar et al., 2003; Sun et al., 2005). Our study was not designed to compare the performance of REPLI-g with regard to degraded DNA, but results pertaining to good quality DNA and WGA amplification could be confirmed.

Zygosity Testing

We genotyped 99 twin pairs with both the SNP and the STR panels, and calculated the sib-ship likelihoods as well as per cent of allele sharing (Appendix A, Supplementary Table 5). When excluding twin pairs with a high proportion of missing markers, 2 pairs out of 98 for the SNP panel, and 2 pairs (numbers 11 and 73) out of 97 for the STR panel, were found to be in conflict with the zygosity assigned from questionnaire. This is in agreement with a previously estimated accuracy of 98% when using validated questionnaire (Lichtenstein et al., 2002). Also, since the genotyping success rate for both of these pairs was 100% for the SNP and the STR panel, missing data can be excluded as a potential explanatory factor for this conflict. Consequently, both panels were equally reliable when it comes to zygosity assignment, but for large-scale studies the slightly higher failure rate, larger work burden, and higher cost for the STR analysis might pose a problem. Here it is important to remember that we were using a noncommercial panel of STRs, and that our results regarding data quality should not be generalized to highly validated and robust commercial kits.

Simulation Studies

We have previously seen that rigorous quality controls are imperative if samples of questionable quality are used for genotyping and concurrent haplotype inference (Pulkkinen et al., 2006). A similar notion has been raised in a study where even a 1% genotyping error was shown to have a big impact on paternity results (Hoffman & Amos 2005). Although SNPs, due to fewer alleles and lower heterozygosity, are not as gravely affected by genotyping error as are STRs (Hoffman & Amos 2005), we still wanted to perform a sensitivity analysis on our SNP panel, in order to investigate how the qualitative results are affected when introducing genotyping error and shifts in the assumed allele frequencies. When using ECLIPSE 2 for calculating the sib-ship likelihoods, one assumes that the allele frequencies of the population from where the twins originate are known. Based on our simulations, we could show that the false positive results of DZ twins were slightly increased when decreasing the minor allele frequency by up to 10% (Table 5). No false positives were observed for the MZ twins. Considering that the SNPs used here are all very polymorphic (Table 1), a decrease in the minor allele frequencies (MAF) of more than 20% might be needed to shift the DZ twin distribution enough to create a substantial overlap with the MZ twin distribution. Such a scenario would be highly unlikely, considering how comparable the allele frequencies are for the different populations examined by the HapMap consortium (Appendix A, Supplementary Table 6). A 1% genotyping error only marginally increased the false positive results for DZ twins. The increase in false positives of MZ twins was only visible when 5 markers were missing in combination with an allele frequency shift of –10%. When no allele frequency shift was present the genotyping error had an impact only when 10 or more markers were missing. The small effect of the genotyping error can be explained by the prior error rate assumed in our analysis. Taken together, the SNP panel seemed to be very robust to moderate changes in the assumed allele frequencies, genotyping errors of up to 1%, and missing data of up to about 20%.

Conclusions

Zygosity testing and general DNA profiling will play an increasing role in research where unambiguous identification of large sets of samples is paramount for performing reliable statistical analyses. We have presented the validation of a set of SNPs that robustly and specifically work as an attractive high throughput and cost efficient option for zygosity testing on a wide range of sample sources. Given the larger number of SNP markers compared to STRs, they provide for some added flexibility when considering the possible presence of somatic mutations and methodological, clerical and genotyping errors that may lead to missing or low quality genotypes for a smaller subset of the markers. The simulations we performed also show the flexibility of the SNP panel as a robust framework for high throughput zygosity screening.

Acknowledgments

We are grateful to Sara Bruce, Kristina Duvefelt and Erik Melén for their valuable comments on the manuscript. This study was supported by the Swedish Research Council.

https://doi.org/10.1375/twin.10.4.604 Published online by Cambridge University Press

References

Barber, A. L., & Foran, D. R. (2006). The utility of whole genome amplification for typing compromised forensic samples. *Journal of Forensic Science*, *51*, 1344–1349.

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics*, *21*, 263–265.

Bergen, A. W., Qi, Y., Haque, K. A., Welch, R. A., & Chanock, S. J. (2005). Effects of DNA mass on multiple displacement whole genome amplification and genotyping performance. *BMC Biotechnology*, *5*, 24.

Boomsma, D., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. *Nature Review. Genetics*, *3*, 872–882.

Dean, F. B., Hosono, S., Fang, L., Wu, X., Faruqi, A. F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S. F., Egholm, M., & Lasken, R. S. (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 5261–5266.

Dixon, L. A., Dobbins, A. E., Pulker, H. K., Butler, J. M., Vallone, P. M., Coble, M. D., Parson, W., Berger, B., Grubwieser, P., Mogensen, H. S., Morling, N., Nielsen, K., Sanchez, J. J., Petkovski, E., Carracedo, A., Sanchez-Diz, P., Ramos-Luis, E., Brion, M., Irwin, J. A., Just, R. S., Loreille, O., Parsons, T. J., Syndercombe-Court, D., Schmitter, H., Stradmann-Bellinghausen, B., Bender, K., & Gill, P. (2005). Analysis of artificially degraded DNA using STRs and SNPs-results of a collaborative European (EDNAP) exercise. *Forensic Science International*, *164*, 33–44.

Guo, S. W., & Thompson, E. A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, *48*, 361–372.

Hannelius, U., Lindgren, C. M., Melen, E., Malmberg, A., von Dobeln, U., & Kere, J. (2005). Phenylketonuria screening registry as a resource for population genetic studies. *Journal of Medical Genetics*, *42*, e60.

Hirschhorn, J. N., Lohmueller, K., Byrne, E., & Hirschhorn, K. (2002). A comprehensive review of genetic association studies. *Genetics in Medicine*, *4*, 45–61.

Hoffman, J. I., & Amos, W. (2005). Microsatellite genotyping errors: Detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology*, *14*, 599–612.

Hosono, S., Faruqi, A. F., Dean, F. B., Du, Y., Sun, Z., Wu, X., Du, J., Kingsmore, S. F., Egholm, M., & Lasken, R. S. (2003). Unbiased whole-genome amplification directly from clinical samples. *Genome Research*, *13*, 954–964.

Ioannidis, J. P., Ntzani, E. E., Trikalinos, T. A., & Contopoulos-Ioannidis, D. G. (2001). Replication validity of genetic association studies. *Nature Genetics*, *29*, 306–309.

Ioannidis, J. P., Trikalinos, T. A., Ntzani, E. E., & Contopoulos-Ioannidis, D. G. (2003). Genetic associations in large versus small studies: An empirical assessment. *Lancet*, *361*, 567–571.

Jackson, R. W., Snieder, H., Davis, H., & Treiber, F. A. (2001). Determination of twin zygosity: a comparison of DNA with various questionnaire indices. *Twin Research*, *4*, 12–18.

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985). Individual-specific 'fingerprints' of human DNA. *Nature*, *316*, 76–9.

Lichtenstein, P., De Faire, U., Floderus, B., Svartengren, M., Svedberg, P., & Pedersen, N. L. (2002). The Swedish Twin Registry: A unique resource for clinical, epidemiological and genetic studies. *Journal of Internal Medicine*, *252*, 184–205.

Liu, K., & Muse, S. V. (2005). PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics*, *21*, 2128–2129.

Lohmueller, K. E., Pearce, C. L., Pike, M., Lander, E. S., & Hirschhorn, J. N. (2003). Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genetics*, *33*, 177–182.

Lovmar, L., Fredriksson, M., Liljedahl, U., Sigurdsson, S., & Syvanen, A. C. (2003). Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DNA. *Nucleic Acids Research*, *31*, e129.

Ooki, S., Yokoyama, Y., & Asaka, A. (2004). Zygosity misclassification of twins at birth in Japan. *Twin Research*, *7*, 228–232.

Peeters, H., Van Gestel, S., Vlietinck, R., Derom, C., & Derom, R. (1998). Validation of a telephone zygosity questionnaire in twins of known zygosity. *Behaviour Genetics*, *28*, 159–163.

Peltonen, L. (2003). GenomEUtwin: A strategy to identify genetic influences on health and disease. *Twin Research*, *6*, 354–360.

Petkovski, E., Keyser-Tracqui, C., Hienne, R., & Ludes, B. (2005). SNPs and MALDI-TOF MS: tools for DNA typing in forensic paternity testing and anthropology. *Journal of Forensic Science*, *50*, 535–541.

Pulkkinen, V., Haataja, R., Hannelius, U., Helve, O., Pitkanen, O. M., Karikoski, R., Rehn, M., Marttila, R., Lindgren, C. M., Hastbacka, J., Andersson, S., Kere, J., Hallman, M., & Laitinen, T. (2006). G protein-coupled receptor for asthma susceptibility associates with respiratory distress syndrome. *Annals of Medicine*, *38*, 357–366.

Reed, T., Plassman, B. L., Tanner, C. M., Dick, D. M., Rinehart, S. A., & Nichols, W. C. (2005). Verification of self-report of zygosity determined via DNA testing in a subset of the NAS-NRC twin registry 40 years

later. *Twin Research and Human Genetics*, *8*, 362–367.

Rylander-Rudqvist, T., Hakansson, N., Tybring, G., & Wolk, A. (2006). Quality and quantity of saliva DNA obtained from the self-administrated oragene method: A pilot study on the cohort of swedish men. *Cancer Epidemiology, Biomarkers and Prevention*, *15*, 1742–1745.

Sieberts, S. K., Wijsman, E. M., & Thompson, E. A. (2002). Relationship inference from trios of individuals, in the presence of typing error. *American Journal of Human Genetics*, *70*, 170–80.

Sun, G., Kaushal, R., Pal, P., Wolujewicz, M., Smelser, D., Cheng, H., Lu, M., Chakraborty, R., Jin, L., & Deka, R. (2005). Whole-genome amplification: Relative efficiencies of the current methods. *Legal Medicine (Tokyo)*, *7*, 279–286.

The International HapMap Project. (2003). *Nature*, *426*, 789–796

Utsuno, H., & Minaguchi, K. (2004). Influence of template DNA degradation on the genotyping of SNPs and STR polymorphisms from forensic materials by PCR. *The Bulletin of Tokyo Dental College*, *45*, 33–46.

Wang, L., Liu, S., Niu, T., & Xu, X. (2005). SNPHunter: A bioinformatic software for single nucleotide polymorphism data acquisition and management. *BMC Bioinformatics*, *6*, 60.

Appendix A

Supplementary Methods and Tables

Microsatellite Genotyping

The PCR was performed in a 384 well format and a total volume of 5 µl. The PCR mix consisted of 0.2 µM forward and reverse primer (Metabion), 0.2 mM dNTP (Roche), 2.5 mM MgCl2 (Qiagen) and 0.2 U Qiagen Hotstart Taq in 1 × PCR buffer was added to 2 ng dried down DNA. The reactions were first optimized in singleplex by running a gradient PCR with 15 minutes of denaturation at 95°C, followed by 40 cycles of 94°C for 30 seconds, 55–65°C for 15 seconds, and 72°C for 30 seconds. A final elongation step of 72°C for 5 minutes ended the program. Next, the primers were run in four multiplexes on 2.5 ng of dried down DNA (Supplementary Table 7) using a touch down PCR temperature profile in order to remedy some unspecific amplification observed after the singleplex run. The temperature profile started with 15 minutes of denaturation at 95°C, followed by 13 cycles of 94°C for 30 seconds, 72°C (–1°C/ cycle) for 15 seconds, and 72°C for 30 seconds. The final 27 cycles were run using a temperature profile of 92°C for 30 seconds, 60°C for 15 seconds, and 72°C for 30 seconds. A final elongation step of 72°C for 5 minutes ended the program. Subsequent PCR:s were run using the same touch down program and the same concentrations of reagents. The DNA template amount was increased to a minimum of 10 ng when genotyping the whole genome amplified samples as well as the twin samples. The final multiplexing scheme is depicted in Supplementary Table 8.

A Beckman Multimek pipetting robot was used for pipetting WGA DNA and gDNA from the twin samples. PCR reagents as well as the gDNA were pipetted using a Hamilton mph96 pipetting station. The reactions for the different markers and the same samples were pooled into 96 well plates for downstream applications using a Beckman Multimek pipetting robot. All primer sequences are available in Supplementary Table 4.

Salt was removed according to manufacturers recommendation by spinning 20 µl of the pooled samples through a Sephadex plate (GE Healthcare).

SNP Genotyping

The PCR was run in a 384 well format using BiometraT100 and Applied Biosystems 9700 thermocyclers. 4 µl of PCR mix containing 0.5 µM PCR primer 1 and 2 accordingly (MetaBion Gmbh, Martinsried, Germany), 0.5 mM dNTP (Roche Diagnostics), 3.5 mM MgCl2 (Qiagen Gmbh, Germany) and 0.5U Qiagen Hotstart Taq for the high-plex pools (> 10-plex), and 0.25 U for the low-plex pool in 1.25 × PCR buffer (Qiagen), was added to a minimum of 7.5 ng (1 µl) of DNA. All PCR reactions used the same temperature profile, with 15 minutes of denaturation at 95°C, followed by 45 cycles of 94°C for 30 seconds, 60°C for 15 seconds, and 72°C for 15 seconds. A final elongation step of 72°C for 5 minutes ended the program. The PCR was set up using a Hamilton mph96 (Hamilton Company, Reno, NV, USA) pipetting station for pipetting of mix, and a Beckman Multimek pipetting robot for dispensing DNA template. All primer sequences are available in Supplementary Table 9.

Unincorporated deoxynucleotides were dephosphorylted and single base extension was performed according to the Sequenom iPLEX protocol. Salt was removed by using an ion exchange resin (Sequenom), after which approximately 10 nl of the samples were spotted onto Maldimatrix containing SpectroCHIPS,

https://doi.org/10.1375/twin.10.4.604 Published online by Cambridge University Press

and analyzed by an Autoflex MassARRAY mass spectrometer (Bruker Daltonics, Billerica, MA, USA). Data were analyzed independently by two persons using the SpectroTyper software (Sequenom Inc.).

DNA extraction

10 ml whole blood was collected in EDTA tubes and DNA extraction was performed on an Autopure LS instrument using Puregene chemistries (Gentra Systems, Inc., Mpl., MN) and the protocol for 5–10 ml whole blood. Two ml saliva was collected using Oragene DNA self-collection kit (DNA Genotek, Canada), and the samples were heated for 1h at 50°C. DNA extraction was performed on the Autopure LS instrument using Puregene chemistries and the protocol for 1 ml cell lysate (J. Dols et. al.). For the filter papers, a few drops of blood were applied onto FTA classic card (Whatman International Ltd.) and the blood spots were left to dry for 1 hour. Genomic DNA for genotyping was extracted using a combination of saponin and chelex-100 (Hannelius et al. 2005), while genomic DNA for REPLI-g amplification was extracted according to the REPLI-g supplementary protocol from February 2005.

DNA Quantification

The DNA from the trios, unrelated Coriell individuals, samples from twins, and the Repli-g amplified DNA from trios and unrelated individuals were quantified using the PicoGreen (Molecular Probes Inc., Eugene, Oregon, USA) assay. The 11 DNA samples from whole blood and saliva, as well as the samples from filter paper used for the WGA reaction, were quantified by both optical density (OD) and PicoGreen. The DNA extracted from filter paper using saponin and chelex-100 was quantified by OliGreen. All corresponding whole genome amplified reactions were quantified by the PicoGreen assay.

The OD quantification was performed in an automated format using Tecan Robot Freedom evo (Tecan Nordic) and GENios spectrophotometer (Tecan Nordic). The OliGreen and PicoGreen assays (Molecular Probes Inc., Eugene, Oregon) were performed according to manufacturer's protocol, using untreated black microtiter well plates (NUNC A/S, Roskilde, Denmark), and fluorescence was measured on a FluoStar Optima (BMG LABTECH GmbH, Germany).

Simulations

The parental genotypes were generated based on allele frequencies and an assumption of HWE. The first twin was consequently generated by assuming a 50% chance of getting either one of two alleles for each marker from the parents. In case of monozygotic twins, a copy of the first twin was then produced, while in case of dizygotic twins, another individual was generated based on the same rules as for the first twin. After having generated a genotype for a specific marker and a specific twin, the genotyping error was introduced, giving each marker and each individual an equal chance of acquiring an error. Missing data were simulated by randomly excluding one marker at a time for all individuals in a simulation set.

https://doi.org/10.1375/twin.10.4.604 Published online by Cambridge University Press

Supplementary Table 1

Sample Id:s of the 14 CEU Trios and 24 Unrelated Individuals Obtained from the Coriell Institute (http://www.coriell.org/) That Were Used for Genotyping

CEU trios	Coriell unrelated
NA06994	NA15029
NA07000	NA15036
NA07029	NA15215
NA07345	NA15223
NA07348	NA15245
NA07357	NA15224
NA10831	NA15236
NA10835	NA15510
NA10839	NA15213
NA10846	NA15221
NA10847	NA15227
NA10851	NA15385
NA10854	NA15590
NA10855	NA15038
NA10857	NA15056
NA10861	NA15072
NA11831	NA15144
NA11832	NA15216
NA11839	NA15226
NA11840	NA15242
NA11994	NA15268
NA11995	NA15324
NA12005	NA15386
NA12006	NA15594
NA12043	
NA12044	
NA12056	
NA12057	
NA12144	
NA12145	
NA12146	
NA12155	
NA12156	
NA12239	
NA12248	
NA12249	
NA12707	
NA12716	
NA12717	
NA12878	
NA12891	
NA12892	

https://doi.org/10.1375/twin.10.4.604 Published online by Cambridge University Press

Supplementary Table 2

SNP Genotyping Success Rates and Concordances

		gDNA Trios + Coriell			WGA DNA Trios and Coriell + gDNA Twins			gDNA Trios versus WGA DNA Trios	gDNA Trios versus HapMap Trios	
chr	ASSAY_ID	SR%	Conc. %	HWE	SR%	Conc. %	HWE	Conc. %	Conc. %	Exclusion criteria
	AMELXY	97%	98%	N/A	98%	100%	N/A	100%	N/A	
2	rs1020636	98%	100%	.232	99%	100%	.009	100%	100%	
1	rs1111366	94%	97%	.548	100%	100%	.512	98%	98%	
5	rs11249784	97%	100%	.732	99%	100%	.344	100%	100%	
3	rs11706962	93%	100%	.061	100%	100%	.253	100%	100%	
X	rs1361861	94%	100%	.685	100%	100%	.672	100%	100%	
8	rs1403294	96%	98%	.700	99%	100%	.547	100%	100%	
3	rs1479530	98%	100%	.360	99%	100%	.102	100%	100%	
12	rs1500098	99%	100%	.279	98%	100%	.188	99%	100%	
11	rs1620329	96%	100%	.4	99%	100%	.176	100%	100%	
X	rs16282	91%	100%	.197	42%	N/A	.536	N/A	100%	low SR%
19	rs1674139	97%	100%	.154	99%	100%	.250	100%	100%	
X	rs17379	92%	97%	.211	98%	100%	.829	98%	97%	
X	rs17407	95%	100%	.142	98%	98%	.446	98%	100%	
9	rs1860665	91%	97%	.776	100%	100%	.507	100%	100%	
1	rs1894697	94%	100%	.031	100%	100%	.152	98%	98%	
13	rs1924609	94%	98%	.585	100%	100%	.754	100%	100%	
X	rs1936827	86%	100%	0	N/A	N/A	N/A	N/A	94%	HWE
2	rs222	98%	100%	.060	99%	100%	.091	100%	100%	
21	rs228043	84%	97%	.364	100%	100%	.101	99%	97%	
1	rs2282739	96%	98%	1	99%	100%	.183	100%	100%	
15	rs2289105	90%	100%	.374	100%	100%	.736	98%	97%	
4	rs230	94%	100%	1	100%	100%	.742	98%	100%	
5	rs2303025	94%	95%	.156	100%	100%	.872	100%	100%	
7	rs234	92%	100%	.552	100%	100%	.422	100%	100%	
9	rs240	94%	100%	.141	100%	100%	.009	100%	100%	
18	rs276922	98%	100%	.263	100%	100%	.190	100%	100%	
3	rs326414	97%	100%	.414	99%	100%	.330	100%	100%	
15	rs3784740	93%	100%	.049	99%	98%	.745	100%	100%	
1	rs4240868	95%	100%	.095	100%	100%	.414	98%	100%	
4	rs4306954	97%	100%	.782	99%	100%	.248	100%	100%	
7	rs4358717	89%	95%	.270	100%	100%	1	98%	97%	
12	rs4763188	98%	100%	.138	99%	100%	.160	100%	100%	
11	rs544021	97%	100%	.381	99%	100%	.422	100%	100%	
14	rs6115	99%	100%	.774	98%	100%	.467	100%	100%	
3	rs6771379	94%	97%	.763	100%	100%	.724	100%	100%	
16	rs710891	83%	97%	.557	94%	100%	.613	98%	100%	
5	rs724784	97%	100%	.261	100%	100%	.681	100%	100%	
20	rs754	96%	100%	.388	99%	100%	.471	100%	100%	
6	rs7747651	54%	100%	0	N/A	N/A	N/A	N/A	96%	HWE
13	rs7994365	97%	100%	.588	99%	100%	.869	100%	100%	
12	rs811	97%	98%	.711	98%	100%	.276	99%	98%	
17	rs820129	97%	100%	.035	99%	100%	.284	100%	100%	
8	rs874746	98%	100%	.154	99%	100%	.028	100%	100%	
11	rs882937	96%	100%	.173	99%	100%	.142	100%	100%	
5	rs889012	97%	100%	.255	99%	100%	.872	100%	100%	
6	rs910170	96%	100%	.101	100%	100%	.243	100%	100%	

(continued over)

https://doi.org/10.1375/twin.10.4.604 Published online by Cambridge University Press

Supplementary Table 2 (continued)

SNP Genotyping Success Rates and Concordances

		gDNA Trios + Coriell			WGA DNA Trios and Coriell + gDNA Twins			gDNA Trios versus WGA DNA Trios	gDNA Trios versus HapMap Trios	
chr	ASSAY_ID	SR%	Conc. %	HWE	SR%	Conc. %	HWE	Conc. %	Conc. %	Exclusion criteria
10	rs9663989	90%	97%	.574	98%	100%	.260	98%	97%	
16	rs9788905	98%	100%	.178	99%	100%	.231	100%	100%	
7	rs997556	95%	100%	1	99%	100%	.097	100%	100%	
Mean		94%	99%		98%	100%		100%	99%	
SD		7%	1%		8%	0%		1%	1%	

Note: The gDNA included DNA from 14 HapMap CEPH (CEU) trios and 24 unrelated individuals obtained from the Coriell Institute while the whole genome amplified DNA and twin samples consisted of the same 14 CEU trios and 24 unrelated individuals as well as gDNA from twins comprising 198 individuals. Both the gDNA and WGA DNA samples were genotyped in duplicates, as were 19 out of the 198 twin samples. HWE for the gDNA samples was calculated using an exact test on genotypes derived from the parents from the 14 CEU trios and for the WGA DNA and twin samples HWE was calculated using the parents from the 14 trios and all 24 unrelated individuals in addition to randomly selected individuals from each twin pair. HWE for X-linked markers was estimated using only female samples. Markers were failed if the genotyping success rate was lower than 80% or if the HWE *p* value was lower than the Bonferroni-adjusted alpha level of .001.

Supplementary Table 3

STR Genotyping Success Rates and Concordances

	gDNA trios			WGA DNA + twins			gDNA versus WGA	
Marker Name	SR%	Conc. %	HWE	SR%	Conc. %	HWE	Conc. %	Exclusion criteria
AMELO	97%	98%	N/A	97%	99%	N/A	98%	
CSF1PO	96%	98%	.036	97%	100%	.397	98%	
D13S317	91%	97%	.277	88%	98%	.869	95%	
D16S539	82%	94%	.795	96%	99%	.062	89%	
D18S51	99%	100%	.868	97%	100%	.015	100%	
D21S11	0%	N/A	N/A	N/A	N/A	N/A	N/A	Failed initial validation
D19S433	99%	100%	.232	87%	97%	0	2%	HWE; global concordance
D2S1338	93%	93%	.527	79%	88%	.041	79%	Low SR%
D3S1358	98%	100%	.770	98%	100%	.281	100%	
D5S818	78%	98%	.215	94%	99%	.426	98%	
D7S820	100%	100%	.234	96%	100%	.617	100%	
D8S1179	0%	N/A	N/A	93%	100%	.854	N/A	Rearranged; failed for gDNA
FGA	98%	100%	.846	95%	99%	.633	95%	
TH01	99%	100%	.917	91%	94%	.849	88%	
TPOX	99%	100%	.550	98%	100%	.103	100%	
vWA	0%	N/A	N/A	92%	99%	.001	N/A	Rearranged; HWE
Total	77%	98%		93%	98%		89%	
SD	39%	2%		5%	3%		25%	

Note: The gDNA included 14 HapMap CEPH (CEU) trios obtained from the Coriell Institute while the whole genome amplified DNA and twin samples consisted of the same 14 CEU trios and additional 24 unrelated individuals of European decent. The twin material consisted of gDNA from 198 individuals. The WGA DNA samples were genotyped in duplicates as were 19 out of the 198 twin samples. All results were analyzed by considering genotypes called independently by two persons. HWE was calculated using an exact test on genotypes derived from the parents from the 14 CEU trios and for the WGA DNA and twin samples the same parents from the 14 trios as well as all 24 Coriell samples in addition to randomly selected individuals from each twin pair. HWE for X-linked markers was estimated using only female samples. Markers were not included if the genotyping success rate was lower than 80% or if the HWE p value was lower than the Bonferroni-adjusted alpha level of .004. Markers D8S1179 and vWA were rearranged following the first genotyping in an attempt to remedy these markers.

https://doi.org/10.1375/twin.10.4.604 Published online by Cambridge University Press

https://doi.org/10.1375/twin.10.4.604 Published online by Cambridge University Press

Supplementary Table 4

SNP Genotyping Success Rates and Concordances for Different DNA Templates

	Replig filter		Replig saliva		Replig wb		Pooled replig	gDNA filter		gDNA saliva		gDNA wb		Pooled gDNA	Pooled all
ASSAY_ID	SR%	Conc.%	SR%	Conc.%	SR%	Conc.%	Conc.%	SR%	Conc.%	SR%	Conc.%	SR%	Conc.%	Conc.%	Conc.%
AMELXY	96%	100%	79%	100%	86%	100%	100%	71%	88%	92%	100%	91%	100%	97%	98%
rs1020636	100%	100%	100%	100%	100%	100%	100%	81%	100%	96%	100%	100%	100%	100%	100%
rs1111366	96%	100%	88%	100%	82%	100%	100%	58%	100%	83%	100%	95%	100%	100%	100%
rs11249784	100%	100%	88%	100%	95%	100%	100%	54%	100%	88%	100%	100%	100%	100%	100%
rs11706962	100%	100%	100%	100%	95%	100%	100%	54%	100%	88%	100%	100%	100%	100%	100%
rs1361861	100%	100%	100%	100%	95%	100%	100%	54%	100%	88%	100%	100%	100%	100%	100%
rs1403294	88%	100%	88%	100%	95%	100%	100%	81%	100%	83%	100%	91%	100%	100%	100%
rs1479530	100%	100%	100%	100%	91%	91%	97%	67%	100%	96%	100%	100%	100%	100%	98%
rs1500098	100%	100%	100%	100%	100%	100%	100%	75%	100%	100%	100%	100%	100%	100%	100%
rs1620329	100%	100%	96%	100%	100%	100%	100%	76%	100%	96%	100%	100%	100%	100%	100%
rs1674139	100%	100%	96%	100%	95%	100%	100%	71%	86%	96%	100%	100%	100%	97%	98%
rs17379	96%	100%	96%	100%	91%	100%	100%	71%	100%	92%	100%	91%	100%	100%	100%
rs17407	92%	100%	88%	100%	100%	100%	100%	67%	100%	92%	100%	100%	100%	100%	100%
rs1860665	100%	100%	100%	100%	100%	100%	100%	71%	100%	92%	100%	100%	100%	100%	100%
rs1894697	100%	100%	96%	100%	100%	100%	100%	54%	100%	92%	100%	100%	100%	100%	100%
rs1924609	100%	100%	100%	100%	86%	100%	100%	83%	91%	92%	100%	100%	100%	97%	98%
rs222	88%	100%	88%	100%	95%	100%	100%	46%	100%	83%	100%	95%	100%	100%	100%
rs228043	96%	100%	100%	100%	95%	100%	100%	83%	100%	88%	100%	95%	100%	100%	100%
rs2282739	100%	100%	100%	100%	100%	100%	100%	81%	100%	92%	100%	100%	100%	100%	100%
rs2289105	83%	100%	96%	100%	86%	100%	100%	58%	100%	88%	100%	95%	100%	100%	100%
rs230	92%	100%	88%	100%	91%	100%	100%	42%	100%	92%	100%	100%	100%	100%	100%
rs2303025	96%	100%	96%	100%	86%	100%	100%	83%	100%	88%	100%	100%	100%	100%	100%
rs234	92%	100%	92%	100%	82%	100%	100%	75%	100%	79%	100%	86%	100%	100%	100%
rs240	100%	100%	100%	100%	100%	100%	100%	54%	80%	92%	100%	100%	100%	96%	98%
rs276922	88%	100%	96%	100%	95%	100%	100%	46%	100%	83%	91%	100%	100%	96%	98%
rs326414	96%	100%	100%	100%	100%	100%	100%	76%	100%	100%	100%	100%	100%	100%	100%
rs3784740	92%	100%	88%	100%	86%	100%	100%	50%	100%	83%	100%	91%	100%	100%	100%
rs4240868	92%	100%	100%	100%	86%	100%	100%	54%	100%	88%	100%	100%	100%	100%	100%
rs4306954	96%	100%	83%	100%	91%	100%	100%	81%	100%	96%	100%	100%	100%	100%	100%
rs4358717	100%	100%	96%	100%	95%	100%	100%	63%	83%	88%	100%	100%	100%	96%	98%
rs4763188	92%	100%	83%	100%	95%	100%	100%	43%	N/A	100%	100%	100%	100%	100%	100%
rs544021	96%	100%	92%	100%	91%	100%	100%	76%	100%	75%	88%	82%	100%	95%	98%

(continued over)

Supplementary Table 4 (continued)

SNP Genotyping Success Rates and Concordances for Different DNA Templates

	Replig filter		Replig saliva		Replig wb		Pooled replig	gDNA filter		gDNA saliva		gDNA wb		Pooled gDNA	Pooled all
ASSAY_ID	SR%	Conc.%	SR%	Conc.%	SR%	Conc.%	Conc.%	SR%	Conc.%	SR%	Conc.%	SR%	Conc.%	Conc.%	Conc.%rs6115
100%	100%	100%	100%	95%	100%	100%	75%	100%	92%	100%	100%	100%	100%	100%	
rs6771379	96%	100%	96%	100%	95%	100%	100%	67%	88%	88%	100%	100%	100%	97%	98%
rs710891	79%	100%	92%	100%	86%	100%	100%	67%	100%	83%	100%	95%	100%	100%	100%
rs724784	88%	100%	100%	100%	95%	100%	100%	71%	100%	88%	100%	100%	100%	100%	100%
rs754	96%	100%	92%	100%	100%	100%	100%	86%	100%	96%	100%	100%	100%	100%	100%
rs7994365	83%	100%	96%	100%	100%	100%	100%	38%	100%	83%	91%	100%	100%	96%	98%
rs811	100%	100%	100%	100%	100%	100%	100%	79%	100%	92%	100%	100%	100%	100%	100%
rs820129	100%	100%	100%	100%	91%	91%	97%	57%	75%	96%	100%	100%	100%	96%	97%
rs874746	96%	100%	100%	100%	100%	100%	100%	62%	100%	92%	100%	100%	100%	100%	100%
rs882937	100%	100%	100%	100%	91%	91%	97%	67%	100%	92%	100%	100%	100%	100%	98%
rs889012	88%	100%	96%	100%	100%	100%	100%	76%	100%	88%	100%	100%	100%	100%	100%
rs910170	100%	100%	100%	100%	95%	100%	100%	79%	100%	88%	100%	100%	100%	100%	100%
rs9663989	88%	100%	88%	100%	82%	100%	100%	57%	100%	83%	100%	100%	100%	100%	100%
rs9788905	100%	100%	100%	100%	100%	100%	100%	81%	100%	92%	100%	100%	100%	100%	100%
rs997556	92%	92%	92%	100%	95%	100%	97%	67%	100%	79%	100%	91%	100%	100%	98%
Total	95%	100%	95%	100%	94%	99%	100%	66%	97%	89%	99%	98%	100%	99%	99%
SD	6%	1%	6%	0%	6%	2%	1%	13%	6%	6%	3%	4%	0%	2%	1%

Note: Genomic DNA (gDNA) and whole genome amplified DNA (WGA DNA) extracted from whole blood, saliva and blood on filter paper obtained from 11 unrelated individuals were genotyped in duplicate using 47 SNP markers.

https://doi.org/10.1375/twin.10.4.604 Published online by Cambridge University Press

Supplementary Table 5

Zygosity Assignments

Pair #	Previously assigned	Twin 1	Twin 2	SNP Results						STR Results					
				Sex twin 1	Sex twin 2	Shared alleles	Missing markers	Per cent shared	Odds (MZ vs DZ)	Sex twin 1	Sex twin 2	Shared alleles	Missing markers	Per cent shared	Odds (MZ vs DZ)
1	1	z00060010	z00060018	M	M	92	0	100%	4.84E+08	M	M	22	0	100%	3.56E+04
2	1	z00060028	z00060055	F	F	92	0	100%	5.12E+08	F	F	22	0	100%	3.78E+04
3	1	z00060049	z00060108	F	F	90	1	100%	1.49E+08	F	F	20	1	100%	1.52E+04
4	1	z00060060	z00060660	F	F	92	0	100%	1.37E+08	F	F	22	0	100%	6.62E+04
5	1	z00060079	z00060436	M	M	92	0	100%	2.78E+08	M	M	22	0	100%	9.56E+04
6	1	z00060106	z00060115	F	F	92	0	100%	3.34E+08	F	F	22	0	100%	2.59E+04
7	1	z00060118	z00060129	F	F	92	0	100%	6.79E+08	F	F	22	0	100%	2.38E+04
8	1	z00060121	z00060711	M	M	92	0	100%	2.74E+08	M	M	22	0	100%	6.69E+04
9	1	z00060127	z00060247	F	F	92	0	100%	1.74E+08	F	F	22	0	100%	7.24E+04
10	1	z00060135	z00071825	M	M	92	0	100%	3.98E+08	M	M	22	0	100%	2.44E+04
11	1	z00060144	z00060190	M	M	76	0	83%	1.42E-15	?	M	14	0	64%	5.43E-15
12	1	z00060151	z00060250	F	F	92	0	100%	3.03E+08	F	F	22	0	100%	3.17E+04
13	1	z00060163	z00060460	M	M	92	0	100%	1.38E+08	M	M	22	0	100%	3.45E+04
14	1	z00060169	z00060193	F	F	92	0	100%	1.36E+08	F	F	22	0	100%	3.82E+04
15	1	z00060184	z00060617	M	M	92	0	100%	2.05E+08	M	M	22	0	100%	2.20E+04
16	1	z00060199	z00060231	M	M	90	1	100%	9.82E+08	M	?	12	5	100%	2.92E+02
17	1	z00060243	z00060256	F	F	92	0	100%	2.58E+08	?	F	12	5	100%	4.46E+02
18	1	z00060261	z00060306	M	M	92	0	100%	6.85E+08	M	M	20	1	100%	9.60E+03
19	1	z00060265	z00060862	F	F	92	0	100%	3.74E+08	F	F	21	0	95%	3.84E+00
20	1	z00060277	z00060942	F	F	92	0	100%	2.43E+08	F	F	22	0	100%	2.67E+04
21	1	z00060280	z00060401	M	M	92	0	100%	2.79E+08	M	M	22	0	100%	3.51E+04
22	1	z00060285	z00060898	M	M	92	0	100%	2.46E+08	M	M	22	0	100%	3.44E+04
23	1	z00060289	z00060292	F	F	90	1	100%	3.36E+08	F	F	22	0	100%	6.81E+04
24	1	z00060297	z00060526	F	F	92	0	100%	3.67E+08	F	F	21	0	95%	2.80E+03
25	1	z00060303	z00060367	F	F	92	0	100%	2.29E+08	F	F	20	1	100%	8.20E+03
26	1	z00060310	z00060321	F	F	92	0	100%	3.43E+08	F	F	22	0	100%	8.11E+04
27	1	z00060351	z00060517	F	F	90	1	100%	2.39E+08	F	F	22	0	100%	3.92E+04
28	1	z00060354	z00060443	F	F	88	2	100%	6.85E+07	F	?	12	5	100%	1.75E+02
29	1	z00060391	z00060581	F	F	92	0	100%	1.01E+08	F	F	22	0	100%	2.96E+04
30	1	z00060394	z00069401	F	F	90	1	100%	1.97E+08	F	F	18	2	100%	3.52E+03
31	1	z00060407	z00060532	M	M	92	0	100%	1.56E+08	M	M	19	1	95%	1.02E+03
32	1	z00060448	z00069415	F	F	92	0	100%	7.39E+08	F	F	22	0	100%	3.32E+04
34	1	z00060479	z00060491	M	M	86	3	100%	7.31E+07	M	M	22	0	100%	3.74E+04
35	1	z00060497	z00060732	M	M	92	0	100%	4.41E+08	M	M	14	4	100%	1.27E+03

(continued over)

Supplementary Table 5

Zygosity Assignments

Pair #	Previously assigned	Twin 1	Twin 2	SNP Results						STR Results					
				Sex twin 1	Sex twin 2	Shared alleles	Missing markers	Per cent shared	Odds (MZ vs DZ)	Sex twin 1	Sex twin 2	Shared alleles	Missing markers	Per cent shared	Odds (MZ vs DZ)
36	1	z00060500	z00060978	F	F	92	0	100%	3.74E+08	F	F	20	1	100%	1.45E+04
37	1	z00060514	z00071334	M	M	92	0	100%	5.04E+08	M	M	22	0	100%	5.76E+04
38	1	z00060541	z00060593	M	M	90	1	100%	1.41E+08	M	M	22	0	100%	4.01E+04
39	1	z00060556	z00060901	M	M	92	0	100%	2.59E+08	M	M	20	1	100%	3.48E+04
40	1	z00060605	z00060921	M	M	92	0	100%	3.24E+08	M	M	20	1	100%	8.29E+03
41	1	z00060729	z00060736	F	F	92	0	100%	3.53E+08	F	F	21	0	95%	2.84E+01
42	1	z00060745	z00069687	M	M	8	41	80%	1.16E-02	M	?	10	5	83%	2.56E-02
43	1	z00060778	z00064680	M	M	90	1	100%	3.69E+08	M	M	20	1	100%	2.02E+04
44	1	z00060847	z00069668	F	F	92	0	100%	4.27E+08	F	F	22	0	100%	6.87E+04
45	1	z00060874	z00069569	M	M	92	0	100%	5.33E+08	M	M	20	1	100%	1.08E+04
46	1	z00068724	z00068730	F	F	92	0	100%	4.54E+08	F	F	14	4	100%	1.21E+03
47	1	z00069466	z00071184	M	M	92	0	100%	1.39E+08	M	M	20	1	100%	1.43E+04
48	1	z00069951	z00069953	F	F	92	0	100%	5.33E+08	F	F	18	2	100%	6.88E+03
49	1	z00069977	z00070002	F	F	92	0	100%	9.72E+08	F	F	22	0	100%	2.50E+04
50	1	z00072181	z00072186	F	F	92	0	100%	3.00E+08	F	F	22	0	100%	2.58E+04
51	2	z00060133	z00060440	F	F	75	0	82%	8.03E-16	F	F	12	0	55%	5.20E-17
52	2	z00060217	z00060325	F	F	75	0	82%	8.04E-15	F	F	17	0	77%	9.99E-08
53	2	z00060315	z00060337	M	M	70	0	76%	3.58E-24	?	M	17	0	77%	5.78E-10
54	2	z00060345	z00069187	M	M	78	0	85%	5.32E-11	M	M	14	1	70%	1.06E-09
55	2	z00060364	z00060535	M	M	75	0	82%	1.06E-17	M	M	15	0	68%	9.69E-15
56	2	z00060413	z00060415	F	F	79	0	86%	1.65E-09	F	F	15	1	75%	8.26E-11
57	2	z00060421	z00060430	F	F	74	0	80%	7.54E-19	F	?	7	5	58%	6.22E-08
58	2	z00060463	z00060538	F	F	77	0	84%	7.15E-15	F	F	16	0	73%	4.56E-08
59	2	z00060466	z00060811	M	M	74	0	80%	3.31E-18	M	M	16	0	73%	2.18E-10
60	2	z00060470	z00072238	F	F	79	0	86%	6.14E-10	F	F	10	5	83%	1.62E-04
61	2	z00060484	z00060502	M	M	67	1	74%	4.62E-24	M	M	17	1	85%	1.18E-02
62	2	z00060506	z00060548	M	M	81	0	88%	7.24E-08	?	M	2	10	100%	2.70E+00
63	2	z00060520	z00069476	F	F	69	1	77%	4.16E-23	F	F	14	0	64%	6.93E-19
64	2	z00060559	z00060596	F	F	76	0	83%	9.50E-15	F	F	14	1	70%	3.02E-07
65	2	z00060723	z00060870	F	F	73	1	81%	3.42E-16	F	F	17	1	85%	6.44E-03
66	2	z00060816	z00071060	F	F	73	0	79%	1.16E-19	F	F	8	5	67%	7.24E-09
67	2	z00060828	z00069200	F	F	79	0	86%	1.59E-09	?	F	14	0	64%	1.58E-14
68	2	z00060868	z00071168	M	M	71	1	79%	5.57E-21	M	M	19	0	86%	5.37E-05
69	2	z00060894	z00069223	M	M	65	0	71%	1.32E-23	N/A	N/A	N/A	N/A	N/A	N/A

(continued over)

https://doi.org/10.1375/twin.10.4.604 Published online by Cambridge University Press

Supplementary Table 5

Zygosity Assignments

				SNP Results						STR Results					
Pair #	Previously assigned	Twin 1	Twin 2	Sex twin 1	Sex twin 2	Shared alleles	Missing markers	Per cent shared	Odds (MZ vs DZ)	Sex twin 1	Sex twin 2	Shared alleles	Missing markers	Per cent shared	Odds (MZ vs DZ)
70	2	z00069220	z00069445	F	F	76	1	84%	1.55E-12	F	F	13	1	65%	2.25E-12
71	2	z00069505	z00069785	F	F	81	0	88%	5.13E-07	F	F	15	1	75%	1.09E-06
72	2	z00069797	z00071786	M	M	70	0	76%	2.06E-21	M	M	14	1	70%	3.83E-10
73	2	z00069894	z00075908	M	M	92	0	100%	8.59E+08	M	M	22	0	100%	4.97E+04
74	2	z00071117	z00071235	M	M	78	0	85%	1.71E-12	M	M	17	1	85%	2.56E-03
75	2	z00071130	z00071178	F	F	75	0	82%	5.94E-16	F	F	14	0	64%	4.89E-18
76	4	z00060067	z00072453	F	M	77	0	84%	1.20E-06	F	M	18	0	82%	3.13E-03
77	4	z00060069	z00060819	M	F	70	0	76%	4.24E-20	M	F	15	1	75%	5.10E-06
78	4	z00060082	z00060201	F	M	78	0	85%	9.69E-08	F	M	16	0	73%	1.62E-06
79	4	z00060088	z00060100	F	M	71	0	77%	1.28E-21	F	M	18	0	82%	8.14E-07
80	4	z00060160	z00060684	M	F	72	0	78%	3.20E-14	M	F	8	4	57%	1.61E-09
81	4	z00060166	z00071765	F	M	77	0	84%	6.52E-14	F	M	17	0	77%	1.31E-10
82	4	z00060214	z00071877	F	M	66	1	73%	8.54E-22	F	M	13	0	59%	1.08E-18
83	4	z00060219	z00069535	M	F	74	0	80%	1.34E-17	M	F	16	0	73%	2.65E-10
84	4	z00060235	z00060853	M	F	74	0	80%	1.57E-14	M	F	16	0	73%	2.41E-10
85	4	z00060300	z00060599	F	M	73	0	79%	3.10E-14	F	M	16	0	73%	4.67E-09
86	4	z00060318	z00069674	M	F	73	0	79%	4.85E-17	M	F	14	0	64%	4.72E-12
87	4	z00060343	z00060703	F	M	76	0	83%	5.21E-12	F	M	19	0	86%	1.53E-02
88	4	z00060388	z00060693	F	M	68	1	76%	1.93E-26	F	M	15	0	68%	1.32E-14
89	4	z00060424	z00069469	F	M	72	0	78%	5.14E-19	F	M	15	0	68%	5.98E-16
90	4	z00060457	z00069683	F	M	67	0	73%	6.25E-29	F	M	13	0	59%	8.76E-17
91	4	z00060607	z00072229	M	F	66	0	72%	1.09E-28	M	F	15	0	68%	6.85E-14
92	4	z00060614	z00076166	M	F	71	0	77%	6.32E-18	M	F	17	0	77%	2.86E-11
93	4	z00060648	z00060982	M	F	37	24	84%	3.34E-07	M	F	15	1	75%	7.56E-09
94	4	z00060804	z00070050	F	M	72	0	78%	1.05E-17	F	M	13	1	65%	2.18E-14
95	4	z00060850	z00068776	F	M	71	0	77%	1.63E-20	F	M	13	0	59%	8.71E-17
96	4	z00060961	z00072223	M	F	77	0	84%	2.66E-10	M	F	11	2	61%	1.55E-13
97	4	z00069185	z00069208	F	M	78	0	85%	2.69E-11	F	M	18	0	82%	7.21E-07
98	4	z00069484	z00069547	F	M	71	0	77%	4.37E-17	F	M	14	0	64%	2.23E-12
99	4	z00071135	z00072328	F	M	69	0	75%	1.09E-22	F	M	13	1	65%	1.45E-11

Note: Previously assigned zygosities are coded as 1) Monozygotic, 2) same sex dizygotic and 4) different sex dizygotic. The odds of the likelihoods of a twin pair being MZ versus DZ are calculated using ECLIPSE2 and a prior genotyping error of 1%. Pairs 62 and 69 failed the STR genotyping, and pair 42 failed the SNP genotyping. Pairs 11 and 73 gave conflicting results when comparing to previously assigned zygosities.

Supplementary Table 6

SNP Marker Information and Allele Frequencies for the Different HapMap Populations and Internally Genotyped Samples

SNP id	chr	Locus	pos	alleles	CEU	CHB	JPT	YRI	Internal
AMELXY	X/Y	*AMEL*	NA	NA	NA	NA	NA	NA	N/A
rs1020636	2	*LTBP1*	33097912	T/C	.62(.49-.74)	.37(.23–.51)	.36(.22–.51)	.85(.76–.94)	.63(.58–.68)
rs1111366	1	*LOC339505*	20463069	C/A	.56(.43-.69)	.50(.35–.65)	.52(.38–.67)	.43(.31–.56)	.60(.55–.66)
rs11249784	5	*LOC441119*	177316388	A/G	.65(.53-.77)	.86(.76–.97)	.85(.74–.96)	.89(.81–.97)	.69(.64–.74)
rs11706962	3	—	196061467	G/A	.62(.50-.75)	.77(.64–.89)	.87(.78–.97)	.97(.94–1.01)	.58(.53–.64)
rs1361861	X	*IL1RAPL2*	103855883	G/A	.52(.39-.64)	.54(.40–.69)	.42(.27–.57)	.55(.42–.68)	.55(.48–.63)
rs1403294	8	—	104414954	T/C	.67(.56–.79)	.82(.71–.93)	.86(.76–.97)	.79(.69–.89)	.73(.67–.78)
rs1479530	3	*CNTN6*	1375419	C/G	.62(.50–.75)	.93(.86–1.01)	.91(.83–.99)	.23(.13–.34)	.64(.58–.69)
rs1500098	12	*IQSEC3*	59354	C/G	.37(.24–.49)	.93(.86–1.01)	.92(.84–1.00)	.73(.62–.84)	.46(.41–.52)
rs1620329	11	*OPCML*	132082653	C/T	.39(.27–.52)	.42(.28–.57)	.45(.31–.60)	.52(.40–.65)	.43(.37–.49)
rs16282	X	*LOC203427*	118362512	C/T	.33(.21–.45)	.37(.23–.52)	.47(.32–.61)	.26(.15–.37)	NA
rs1674139	19	*AP2A1*	54963314	T/C	.38(.26–.51)	.45(.31–.60)	.42(.27–.57)	.62(.50–.75)	.43(.37–.49)
rs17379	X	*HS6ST2*	131802447	A/G	.57(.45–.70)	.55(.40–.69)	.82(.70–.93)	.26(.15–.37)	.51(.43–.59)
rs17407	X	—	140808205	A/G	.62(.50–.75)	1.00(1.00–1.00)	1.00(1.00–1.00)	.59(.47–.72)	.67(.60–.75)
rs1860665	9	—	119501589	T/G	.46(.33–.58)	.66(.52–.79)	.72(.58–.85)	.43(.31–.56)	.51(.45–.57)
rs1894697	1	*F5*	166258259	G/C	.50(.37–.63)	.10(.01–.19)	.15(.04–.25)	.17(.07–.26)	.49(.44–.55)
rs1924609	13	*ATP7B*	51410519	C/T	.49(.37–.62)	.64(.50–.78)	.59(.45–.74)	.43(.31–.56)	.49(.44–.55)
rs1936827	X	–	125063418	G/A	.63(.51–.76)	.47(.32–.61)	.50(.35–.65)	.15(.06–.24)	NA
rs222	2	*INPP4A*	98606390	T/C	.76(.65–.87)	.72(.58–.85)	.81(.69–.92)	.45(.32–.58)	.79(.74–.83)
rs228043	21	*SLC37A1*	42823507	A/G	.51(.38–.64)	.75(.62–.88)	.80(.68–.91)	.43(.31–.56)	.53(.47–.59)
rs2282739	1	*HSD11B1*	206271396	C/T	.74(.62–.85)	.53(.39–.68)	.43(.29–.58)	.23(.13–.34)	.70(.65–.76)
rs2289105	15	*CYP19A1*	49294800	T/C	.41(.28–.53)	.39(.25–.53)	.58(.43–.72)	.85(.76–.94)	.52(.46–.58)
rs230	4	—	173271203	A/G	.39(.27–.52)	.72(.59–.85)	.59(.45–.74)	.81(.71–.91)	.47(.41–.52)
rs2303025	5	*ANXA6*	150483868	T/C	.56(.43–.69)	.43(.29–.58)	.37(.23–.52)	.40(.28–.52)	.59(.54–.65)
rs234	7	—	105155086	T/C	.56(.43–.68)	.30(.17–.43)	.33(.19–.47)	.29(.18–.41)	.53(.47–.58)
rs240	9	—	109558585	C/G	.55(.42–.68)	.61(.47–.75)	.61(.47–.76)	.52(.39–.64)	.63(.57–.68)
rs276922	18	*DSC3*	26858793	A/C	.57(.44–.69)	.20(.08–.32)	.08(.00–.16)	.53(.41–.66)	.52(.46–.57)
rs326414	3	—	7878040	A/G	.55(.42–.68)	.61(.47–.75)	.62(.48–.76)	.33(.21–.45)	.52(.47–.58)
rs3784740	15	*ST8SIA2*	90761479	T/C	.55(.42–.68)	.19(.07–.30)	.30(.16–.43)	.55(.42–.68)	.55(.50–.61)
rs4240868	1	—	149794672	G/A	.52(.39–.64)	.38(.24–.52)	.32(.18–.46)	.54(.42–.67)	.46(.40–.51)
rs4306954	4	—	105660759	T/C	.53(.40–.65)	.76(.63–.88)	.80(.68–.91)	.15(.06–.24)	.49(.43–.55)
rs4358717	7	—	140396256	T/C	.57(.44–.70)	.13(.03–.23)	.17(.06–.29)	.33(.21–.45)	.56(.51–.62)
rs4763188	12	*PPM1H*	61333020	T/A	.59(.47–.72)	.60(.46–.74)	.49(.34–.64)	.60(.48–.73)	.65(.60–.70)

(continued over)

Supplementary Table 6 (continued)

SNP Marker Information and Allele Frequencies for the Different HapMap Populations and Internally Genotyped Samples

SNP id	chr	Locus	pos	alleles	CEU	CHB	JPT	YRI	Internal
rs544021	11	*ACTN3/ZDHHC24*	66072237	C/G	.57(.44–.69)	.60(.46–.75)	.64(.49–.78)	.32(.21–.44)	.55(.49–.61)
rs6115	14	*SERPINA5*	94123643	G/A	.38(.26–.51)	.57(.42–.71)	.64(.49–.78)	.02(–.01–.06)	.35(.29–.40)
rs6771379	3	—	5590030	C/T	.32(.21–.44)	.13(.02–.23)	.09(.00–.18)	.54(.40–.67)	.36(.30–.41)
rs710891	16	*TRAP1*	3650098	C/T	.65(.53–.77)	.57(.42–.71)	.42(.27–.57)	.10(.02–.18)	.60(.54–.66)
rs724784	5	*MAN2A1*	109081234	A/C	.66(.54–.78)	.76(.63–.88)	.83(.72–.94)	.30(.18–.42)	.61(.56–.67)
rs754	20	—	58102499	A/G	.92(.85–.99)	.63(.49–.77)	.67(.53–.81)	.33(.21–.45)	.87(.84–.91)
rs7747651	6	*EFHC1*	52433769	C/A	.63(.51–.76)	.77(.64–.89)	.77(.65–.90)	.48(.36–.61)	NA
rs7994365	13	*AKAP11*	41765954	A/G	.42(.30–.55)	.26(.13–.38)	.42(.27–.57)	.44(.32–.57)	.45(.39–.50)
rs811	12	—	89278405	G/A	.35(.23–.47)	.23(.11–.36)	.12(.03–.22)	.01(–.01–.03)	.35(.30–.41)
rs820129	17	*SAP30BP*	71179702	A/G	.67(.55–.79)	.39(.24–.53)	.50(.35–.65)	.34(.22–.46)	.65(.60–.70)
rs874746	8	*KCNK9*	140722490	C/A	.45(.32–.58)	.21(.09–.33)	.15(.04–.25)	.47(.35–.60)	.41(.36–.47)
rs882937	11	*PANX1*	93548520	A/G	.44(.32–.57)	.55(.40–.69)	.40(.26–.54)	.55(.42–.68)	.44(.39–.50)
rs889012	5	*TRPC7*	135582615	C/A	.48(.36–.61)	.62(.48–.76)	.58(.43–.73)	.60(.48–.72)	.55(.50–.61)
rs910170	6	*GLP1R*	39140393	T/C	.44(.32–.57)	.30(.17–.43)	.49(.34–.64)	.37(.25–.50)	.45(.40–.51)
rs9663989	10	*TLL2*	98225107	A/C	.39(.27–.52)	.56(.41–.70)	.60(.46–.75)	.57(.45–.70)	.44(.39–.50)
rs9788905	16	—	65000382	T/C	.61(.48–.73)	.88(.78–.97)	.90(.81–.99)	.78(.68–.89)	.63(.58–.68)
rs997556	7	—	51738954	T/C	.69(.57–.81)	.47(.32–.61)	.40(.25–.54)	.53(.41–.66)	.59(.54–.65)

Note: The allele frequencies including 95% confidence intervals for the following HapMap populations are represented; 30 trios from Utah Residents with Northern and Western European Ancestry (CEU), 45 unrelated Han Chinese in Beijing, China (CHB), 45 unrelated Japanese in Tokyo, Japan (JPT), 30 trios from Yoruba in Ibadan, Nigeria (YRI), and the internal results stemming from 15 CEU trios, 24 unrelated individuals from the Coriell Institute and 99 randomly chosen individuals from each genotyped twin pair. For X-chromosomal markers the allele frequencies are calculated using only female individuals. The reference SNP id:s, chromosomal positions and gene symbols are presented according to NCBI build 35 and the allele frequencies represent nonredundant HapMap data released on the July 20, 2006. All frequencies are presented in relation to the HapMap reference allele. The reference allelle allele is presented as the first allele in the alleles-column.

Supplementary Table 7

STR Primer Sequences and Fluorescent Tags Used in the First Genotyping Round

	Multiplex pool	PCR primer 1	PCR primer 2	Fluorescent tag
AMELO	W1	GACCAGAATATGAGACAGGAACTG	TTGCTAAGTTAAGTGATTGTAAGCA	NED
CSF1PO	W1	AACCTGAGTCTGCCAAGGACTAGC	TTCCACACACCACTGGCCATCTTC	FAM
D13S317	W1	ACAGAAGTCTGGGATGTGGA	GCCCAAAAAGACAGACAGAA	FAM
D18S51	W1	GAGCCATGTTCATGCCACTG	CAAACCCGACTACCAGCAAC	VIC
D21S11	W1	GTGAGTCAATTCCCCAAG	GTTGTATTAGTCAATGTTCTCC	VIC
D3S1358	W1	ACTGCAGTCCAATCTGGGT	ATGAAATCAACAGAGGCTTG	VIC
D8S1179	W1	TTTTTGTATTTCATGTGTACATTCG	CGTAGCTATAATTAGTTCATTTTCA	NED
FGA	W1	GCCCCATAGGTTTTGAACTCA	TGATTTGTCTGTAATTGCCAGC	NED
D16S539	W2	GATCCCAAGCTCTTCCTCTT	ACGTTTGTGTGTGCATCTGT	FAM
D19S433	W2	CCTGGGCAACAGAATAAGAT	TAGGTTTTTAAGGAACAGGTGG	FAM
D5S818	W2	GGGTGATTTTCCTCTTTGGT	TGATTCCAATCATAGCCACA	FAM
TH01	W2	GTGGGCTGAAAAGCTCCCGATTAT	GTGATTCCCATTGGCCTGTTCCTC	VIC
TPOX	W2	CACTAGCACCCAGAACCGTC	CCTTGTCAGCGTTTATTTGCC	NED
vWA	W2	CCCTAGTGGATGATAAGAATAATC	GGACAGATGATAAATACATAGGATGGATGG	NED
D2S1338	W3	CCAGTGGATTTGGAAACAGA	ACCTAGCATGGTACCTGCAG	VIC
D7S820	W4	TGTCATAGTTTAGAACGAACTAACG	CTGAGGTATCAAAAACTCAGAGG	FAM

Note: The colors of fluorescent tags are yellow (NED) blue (FAM) and green (VIC)

Supplementary Table 8

STR Primer Sequences and Fluorescent Tags Used in the Second Genotyping Round

	Multiplex pool	PCR primer 1	PCR primer 2	Fluorescent tag
AMELO	W1	GACCAGAATATGAGACAGGAACTG	TTGCTAAGTTAAGTGATTGTAAGCA	NED
CSF1PO	W1	AACCTGAGTCTGCCAAGGACTAGC	TTCCACACACCACTGGCCATCTTC	FAM
D13S317	W1	ACAGAAGTCTGGGATGTGGA	GCCCAAAAAGACAGACAGAA	FAM
D18S51	W1	GAGCCATGTTCATGCCACTG	CAAACCCGACTACCAGCAAC	VIC
D3S1358	W1	ACTGCAGTCCAATCTGGGT	ATGAAATCAACAGAGGCTTG	VIC
FGA	W1	GCCCCATAGGTTTTGAACTCA	TGATTTGTCTGTAATTGCCAGC	NED
D16S539	W2	GATCCCAAGCTCTTCCTCTT	ACGTTTGTGTGTGCATCTGT	FAM
D19S433	W2	CCTGGGCAACAGAATAAGAT	TAGGTTTTTAAGGAACAGGTGG	FAM
D5S818	W2	GGGTGATTTTCCTCTTTGGT	TGATTCCAATCATAGCCACA	FAM
TH01	W2	GTGGGCTGAAAAGCTCCCGATTAT	GTGATTCCCATTGGCCTGTTCCTC	VIC
TPOX	W2	CACTAGCACCCAGAACCGTC	CCTTGTCAGCGTTTATTTGCC	NED
D8S1179	W3	TTTTTGTATTTCATGTGTACATTCG	CGTAGCTATAATTAGTTCATTTTCA	NED
D2S1338	W3	CCAGTGGATTTGGAAACAGA	ACCTAGCATGGTACCTGCAG	VIC
D7S820	W4	TGTCATAGTTTAGAACGAACTAACG	CTGAGGTATCAAAAACTCAGAGG	FAM
vWA	W4	CCCTAGTGGATGATAAGAATAATC	GGACAGATGATAAATACATAGGATGGATGG	NED

Note: Marker D2S11 was failed after the first round of genotyping. The colors of fluorescent tags are yellow (NED), blue (FAM) and green (VIC)

https://doi.org/10.1375/twin.10.4.604 Published online by Cambridge University Press

https://doi.org/10.1375/twin.10.4.604 Published online by Cambridge University Press

Supplementary Table 9

SNP Primer Sequences

Multiplex pool	SNP name	PCR primer 1	PCR primer 2	Extension primerpool
W1	rs882937	ACGTTGGATGAATTCATAGCTGGCTGTGGG	ACGTTGGATGCTCTGTGTCATGCACTTATG	CTAATGGGCACCCTTC
W1	rs1924609	ACGTTGGATGACACCTGCCTGTCACATTCG	ACGTTGGATGCCTTTTCTGGCTAGCACTTG	ATTCAGCAGGCACTTA
W1	rs4358717	ACGTTGGATGGCAGAAGTTAACGGTTGGAG	ACGTTGGATGGTGTGTGGTTTTGCAGAGAG	GAGAGGGCCTCTGGGA
W1	rs234	ACGTTGGATGTATTCAAGGAGTGTCCCTGG	ACGTTGGATGCACCCATCACTCACAGTTAC	TGTGGCAGAGACTGAAT
W1	rs2303025	ACGTTGGATGTCAAGTTTCCACCCTCCTAG	ACGTTGGATGCTTCTGCCAAGTTCCTTATG	TCCTTATGAACCCTCTCC
W1	rs1894697	ACGTTGGATGAGTCAAATCACCAGAACCGC	ACGTTGGATGTTCCCTTAGGCAGAGCTTCG	GCAGAGCTTCGACCACTA
W1	rs1111366	ACGTTGGATGGCCTCCCACATTATGTAGTC	ACGTTGGATGAAGAGAGTACCTAGTCAGGC	CATCCTTGTTGGGTTTAG
W1	rs240	ACGTTGGATGTCAACCTCCGACTTTCACAG	ACGTTGGATGCCATCAAATGCCTTTCTCCC	CAGAAGCAATTACAGGAAG
W1	rs230	ACGTTGGATGGGAGATAAGTATATGGTAGGC	ACGTTGGATGATGTCCTCCCAAGTGTGATG	GTGTGATGCTCTACCCTACT
W1	rs16282	ACGTTGGATGGTATAAGAGTGGGCTGGAAT	ACGTTGGATGGATTCTACATTACATTGTAT	TGTATAAGTCTTGACCTACC
W1	rs11706962	ACGTTGGATGAGGCTGCTTCCACTTATGGG	ACGTTGGATGGTTAGGAAATGCCTGGCACC	TGTCTTTTTCCTCCTCTCTTG
W1	rs276922	ACGTTGGATGGACAAATTGGGAATTGCTGC	ACGTTGGATGCAGAAAATTATGCTGGAGAG	GGAGAGAACAATTAAACTCAC
W1	rs4240868	ACGTTGGATGCTGCACAGTATAGCATTGGC	ACGTTGGATGTGCCAGTTTGGCAACTACTC	AATTTCTACAGTCTACTGTTTC
W1	rs11249784	ACGTTGGATGCTTCAGAATTTGTTTTCCTCC	ACGTTGGATGCATTTTGCTGACCAAGAACG	TTTTGTCTCCAGTGTTCAAATA
W1	rs1361861	ACGTTGGATGAGAAGATAGAGCTGAGAGGC	ACGTTGGATGTTTACAGGTGGAGAGGGATG	GAGAGGGATGTATACACTGGAC
W1	rs910170	ACGTTGGATGTAAGTTCCCCTACAGAGAGG	ACGTTGGATGTGATGTGAGGCTGATAGAGG	TGATAGAGGCCACTGCTTACTTG
W1	rs3784740	ACGTTGGATGACTAGAACCCATCAGGAACC	ACGTTGGATGTTACCACTCAGGAAGCCTTG	ACTCAGGAAGCCTTGGTGTCAGA
W1	rs1860665	ACGTTGGATGGGGAGATTGTGAAGATAGGG	ACGTTGGATGTCAGCCACTTACTTGCATGG	CCACTTACTTGCATGGCCATACTT
W1	rs6771379	ACGTTGGATGGTCCCCTTTGAAGAATCCAC	ACGTTGGATGTCTCAACTGCTTTCTCTACC	GCTTTCTCTACCTTGTCCTTTACTC
W1	rs7994365	ACGTTGGATGCTTCAAAACCCTATGCCAGC	ACGTTGGATGGCACACCCTAAAATGGAATAG	CCTAAAATGGAATAGAAATTCCATC
W1	rs997556	ACGTTGGATGTGAGCTGCACAAAATGGAGG	ACGTTGGATGTGGAGAGTCACAAAATGGCC	GAGAGTCACAAAATGGCCCTTATTA
W1	rs724784	ACGTTGGATGCTGAAGTCCTTCCATGATGC	ACGTTGGATGCTCCTTAATCTTAGCCTGTAC	TCCTTAATCTTAGCCTGTACCTTTTA
W1	rs2289105	ACGTTGGATGAAAAAATGAGGGGAGGTGAC	ACGTTGGATGGTCCACAGTCAATCACAGAG	ATGATTTCATTTTGTTGAGGTTGTTG
W1	rs222	ACGTTGGATGGGAAAAAAAAATCAGAGGAGAG	ACGTTGGATGGCTTCTGACCCTTTTCTGTG	ATAAATGTAACATTAGACCTCTCACTA
W1	rs228043	ACGTTGGATGTCATGAGTGGAGCATTTGCG	ACGTTGGATGACAGGTGCCACGTGAAATGC	GTGCCACGTGAAATGCATTATACACAC
W2	rs326414	ACGTTGGATGGCATTTTTGAGGGGTCTAAG	ACGTTGGATGTATGGAGGACTTGATGAACC	ACCCAGATTCTGGTGC
W2	rs754	ACGTTGGATGCATGGAGACATTCATTAGGC	ACGTTGGATGAATCATAATGCCTGCCCCTC	AGGAAGTTCTGAGGGT
W2	rs874746	ACGTTGGATGGGTTTCTTAGGACCAAATGAG	ACGTTGGATGGATAGTGAACCTGCCTCACC	TGCCAGACCGCTTCTTG
W2	rs544021	ACGTTGGATGTAGATGACAGGAGATGACCC	ACGTTGGATGCTAAGGGGAGGCCTGTCAC	GAGTCAGCGGAGCCTCTC
W2	rs9663989	ACGTTGGATGAACCCAGGCATAAGGTCTTG	ACGTTGGATGAGAGTGCAGGTGGTTATTCC	CAAGTAGGCTGGAGTCTT
W2	rs1403294	ACGTTGGATGTTACCTGGTGCTTGAGCTAC	ACGTTGGATGGCTTCAAGTTGTCTGGGTAG	GTTGTCTGGGTAGATAGAA
W2	rs1674139	ACGTTGGATGGAAGAGTTGTCGTGACGATG	ACGTTGGATGCCAATCACCCACAGCCATTC	CCGCGAGCTGACTTGCCGCC
W2	rs820129	ACGTTGGATGCTGTGTGGACATTTTCTGGG	ACGTTGGATGGGCTGGAAATCCAATGTGTC	GGATTTGTTCATTCCTGAGT
W2	rs889012	ACGTTGGATGACCAGCATGGAAATCATGGC	ACGTTGGATGGACTAATGGGAGCCAGTTTC	ACATCTCATTTCCTAAACCAT
W2	rs1479530	ACGTTGGATGCCCTCATCCCAACATGAAAC	ACGTTGGATGGAAATTTTTATTGTGCCTGTC	TTCTTGTGCCTGTCAATAGGT
W2	rs4763188	ACGTTGGATGAGTGGTAAGCTCACCTAGTC	ACGTTGGATGTAAAGCTATCAGCCAAAATC	GGAGCAGCCAAAATCAAAGGT
W2	rs1936827	ACGTTGGATGGTGTGCTACTGTGAGTCTAT	ACGTTGGATGGTGTACATGATCATTCAATC	CAATGATCATTCAATCTCCATC
W2	rs710891	ACGTTGGATGCCTCACATTTCTAGTCTACG	ACGTTGGATGTAGGGTAACTAGGTGACCTC	CACCTCCACATTAGCATATTAA

(continued over)

Supplementary Table 9 (continued)

SNP Primer Sequences

Multiplex pool	SNP name	PCR primer 1	PCR primer 2	Extension primer
W2	rs2282739	ACGTTGGATGTTCAGACATGCCAATGAGCC	ACGTTGGATGCCTTCTATCTCAGAACCACC	AACGCCTAGAGTGCTTGTTTACA
W2	rs9788905	ACGTTGGATGAGTCAGCAGTCCTGGATGAG	ACGTTGGATGAAAATTGTGACTCAGGGAGG	AAGCAACTAACTCTGCTTGTAGG
W2	rs1020636	ACGTTGGATGAAGGCTGGTCTGGCAATCTG	ACGTTGGATGACAGTGAAGTACCCAAGTCC	AGTCCACACACAAATTATTCTCAT
W2	rs4306954	ACGTTGGATGATGCACAGCTCTTTCACTAG	ACGTTGGATGGTGGCAAGTAGAAACATTCAC	AGTAGAAACATTCACATATGAATTA
W2	rs1620329	ACGTTGGATGATCTGCCCACTGACAATGAG	ACGTTGGATGTCAGCCATGTGCCTGGATAA	GTAATAATGTGTCTATACTAGAAGT
W2	rs7747651	ACGTTGGATGCAATATTGAATTTGTGAGAG	ACGTTGGATGCCTATCACCCTTTCTTATGC	TCTCTTATGCTATTACTGTAATACAT
W2	rs17407	ACGTTGGATGGGAAAGAATCACGGGAATTAC	ACGTTGGATGTTTCTGGCTTCATCTACCTC	AGTAGAATTTGATTTCTACTCCTAAAA
W2	rs17379	ACGTTGGATGGAGCAAGATTCTTTGGCCTC	ACGTTGGATGTTCTACCCCCTTGTAGACTG	AACACTTGTAGACTGCATTTTTGGTAC
W3	rs6115	ACGTTGGATGTCATGGAGATGCTCACAGGG	ACGTTGGATGGCAGAAGGGACTTTACCT	GCTGCCCCCAGCCAGA
W3	rs1500098	ACGTTGGATGCCAGCTATATGTACATTTAAGC	ACGTTGGATGTTAGGGAACAAAACTCAAGG	AACTCAAGGTGCAGAA
W3	rs811	ACGTTGGATGGAGGAAGAAGAGGGAGTAAG	ACGTTGGATGAGGTACCAGAGTCATGGAGG	GTCATGGAGGGTCATTA
W3	AMELXY	ACGTTGGATGACCAAATCATCCCCGTGGTG	ACGTTGGATGAGCTGGCACCACTGGGATGT	TGGGATGTGGTGATGAG

Note: PCR and extension primer sequences for the 50 SNP markers. The multiplexes, designed using Sequenoms Assay Design v3.4, are presented as W1, W2 and W3