Twins, Tissue, and Time: An Assessment of SNPs and CNVs

Paul Scheet, Erik A. Ehli, Xiangjun Xiao, Catharina E. M. van Beijsterveldt, Abdel Abdellaoui, Robert R. Althoff, Jouke Jan Hottenga, Gonneke Willemsen, Kelly A. Nelson, Patricia E. Huizenga, Yueshan Hu, Christopher I. Amos, Meike Bartels, Maria M Groen-Blokhuis, Eco JC de Geus, James J. Hudziak, Gareth E. Davies and Dorret I. Boomsma

How to cite this article:
Paul Scheet, Erik A. Ehli, Xiangjun Xiao, Catharina E. M. van Beijsterveldt, Abdel Abdellaoui, Robert R. Althoff, Jouke Jan Hottenga, Gonneke Willemsen, Kelly A. Nelson, Patricia E. Huizenga, Yueshan Hu, Christopher I. Amos, Meike Bartels, Maria M Groen-Blokhuis, Eco JC de Geus, James J. Hudziak, Gareth E. Davies and Dorret I. Boomsma (2012). Twins, Tissue, and Time: An Assessment of SNPs and CNVs. Twin Research and Human Genetics, 15, pp 737-745 doi:10.1017/thg.2012.61

Request Permissions : Click here

Twins, Tissue, and Time: An Assessment of SNPs and CNVs

Paul Scheet,[1*] Erik A. Ehli,[2,3*] Xiangjun Xiao,[1] Catharina E. M. van Beijsterveldt,[4] Abdel Abdellaoui,[4] Robert R. Althoff,[5] Jouke Jan Hottenga,[4] Gonneke Willemsen,[4] Kelly A. Nelson,[2] Patricia E. Huizenga,[2] Yueshan Hu,[2,3] Christopher I. Amos,[6] Meike Bartels,[4] Maria M Groen-Blokhuis,[4] Eco JC de Geus,[4] James J. Hudziak,[5] Gareth E. Davies,[2,3] and Dorret I. Boomsma[4]

[1]*Department of Epidemiology, University of Texas MD Anderson Cancer Center, Houston, TX, USA*

[2]*Avera Institute for Human Genetics, Avera McKennan Hospital and University Health Center, Sioux Falls, SD, USA*

[3]*Department of Psychiatry, University of South Dakota, Sanford School of Medicine, Sioux Falls, SD, USA*

[4]*Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands*

[5]*Department of Psychiatry, University of Vermont, College of Medicine, Burlington, VT, USA*

[6]*Department of Genetics, University of Texas MD Anderson Cancer Center, Houston, TX, USA*

These authors contributed equally to this work.

With the desire to assess genetic variation across the lifespan in large-scale collaborative projects, one question is whether inference of copy number (CN) is sensitive to the source of material for deoxyribonucleic acid (DNA) analysis (e.g., blood and buccal) and another question is whether CN is stable as individuals age. Here, we address these questions by applying Affymetrix 6.0 single nucleotide polymorphism (SNP) micro-arrays to 1,472 DNA samples from 710 individuals from the Netherlands Twin Register, including twin and non-twin individuals (372 with buccal and blood derived DNA and 388 with longitudinal data). Similar concordance for CN and genotype inference between samples from the same individual [or from the monozygotic (MZ) co-twins] was found for blood and buccal tissues. There was a small but statistically significant decrease in across-tissue concordance compared with concordance of samples from the same tissue type. No temporal effect was seen on CN variation from the 388 individuals sampled at two time points ranging from 1 to 12 years apart. The majority of our individuals were sampled at age younger than 20 years. Genotype concordance was very high ($R^2 > 99\%$) between co-twins from 43 MZ pairs. For 75 dizygotic (DZ) pairs, R^2 was $\approx 65\%$. CN estimates were highly consistent between co-twins from MZ pairs for both deletions ($R^2 \approx 90\%$) and duplications ($R^2 \approx 86\%$). For DZ, these were similar for within-individual comparisons, but naturally lower between co-twins ($R^2 \approx 50$–60%). These results suggest that DNA from buccal samples perform as well as DNA from blood samples on the current generation of micro-array technologies.

■ **Keywords:** copy number variation, buccal-derived DNA, twins, somatic variation

Recent interest in the role that copy number variation (CNV) plays in common diseases has generated a remarkable increase in research using a variety of tissue types to explore the etiopathology of CNVs. Large-scale epidemiologic research projects of common phenotypes in children may benefit from being able to use buccal samples as a source of deoxyribonucleic acid (DNA) for genome-wide association (GWA) studies of single nucleotide polymorphisms (SNPs) and CNVs. One obvious benefit of using buccal samples in young children is the lack of the need to use painful needle sticks (Dlugos et al., 2005). Other advantages include cost and the ability to collect samples via a variety of methods, for example, through the mail. However, although there is little debate about the utility of buccal DNA for SNP analyses, there has been some uncertainty about the use of buccal DNA for CNV research. Concerns focus on measurement error, bacterial contamination, and other sources of variance that may be introduced into analyses because of the environment in which the buccal cell exists (e.g., oral

TABLE 1

Descriptive Participant Characteristics for Tissue and Temporal Contrasts

	No. samples	Median age at sampling	Tissue type Blood	Tissue type Buccal	Female (%)
Twin, tissue contrasts					
MZ sibships (44)	86	18 (18, 19)	86	88	53
DZ sibships (75)	150	18 (16, 30)	150	150	52
Single twin	65	18 (4, 31)	65	65	62
Single non-twin	71	20 (12, 35)	71	71	61
Unique individuals tissue	**372**	18 (4, 35)	**372**	374	**56**
Longitudinal contrasts					
1-year interval	2	18.5 (18, 19)	0	2	100
2 years	257	10 (8, 46)	0	257	52
3 years	8	31 (15, 52)	0	8	50
7 years	65	9 (5, 13)	0	65	38
8 years	14	8.5 (4, 13)	0	14	57
10 years	1	13 (8, 18)	1	0	0
11 years	19	12.5 (7, 18)	19	0	63
12 years	23	12.5 (6, 19)	23	0	48
Total longitudinal contrasts	**389**	10 (4, 52)	**43**	346	**50**
Both blood and buccal longitudinal sample	1				
Unique longitudinal contrasts	**388**				
Tissue and longitudinal overlap	50				
Unique individuals	**710**				**53**

Note: The total number of individuals involved in this study is broken down by assessments of consistency across tissue type as well as longitudinal. Median age at sampling is reported (with range given in parentheses). Fifty individuals had samples that were used in both groups of contrasts; thus, the total number of unique individuals is less than the sums of the individuals involved in the tissue and longitudinal assessments.

flora and fauna). Efforts to address these concerns about measurement error and technical repeatability focused on genotype estimation (e.g., Meulenbelt et al., 1995; Min et al., 2006; Woo et al., 2007) or have been conducted on non-human populations (Rincon et al., 2011). For CNVs, much less is known about the effect of source material (Dellinger, 2010). Multiple algorithms have emerged for the detection of CNV, and performance of these algorithms has been compared (Pinto et al., 2011), but comparisons have not included an assessment across tissue.

In this study, we analyzed DNA from longitudinal blood and buccal sources from a large sample (710 individuals; 335 male and 375 female) of individuals, including 43 monozygotic (MZ; 23 female) and 75 dizygotic (DZ; 26 both female and 23 both male) twin pairs, their parents and siblings (see Table 1). The participants are registered with the Netherlands Twin Register (NTR; Bartels et al., 2007; Boomsma et al., 2006) and DNA samples have been collected in multiple projects, both from blood and buccal material. Previously, the longitudinal samples were used to study the across-tissue and the temporal stability of DNA methylation (Talens et al., 2010). In this article, we use the availability of the longitudinal samples from the same and from different tissues to obtain a technical evaluation and confirmation of CN inference from the Affymetrix 6.0 DNA micro-array technology with CN calls made using Birdsuite v.2 (Korn et al., 2008). Finally, we evaluate whether blood is an acceptable DNA source in studies that include twins, as the utility of blood samples that come from twins has been questioned because genomic DNA from MZ as well as DZ twins (e.g., see van Dijk et al., 1996) derived from blood samples may reflect a shared circulation prenatally (Erlich, 2011).

Biological materials were collected over a period of 20 years in different NTR projects, but SNP and CN data were generated in a single laboratory, as laboratory settings have also been implicated as a source of variation (Pinto et al., 2011). All DNA samples collected from different tissues and time points of the same individual, and all DNA from related individuals were run on the same plate.

To assess the consistency of CN inference across DNA source material, blood- and buccal-derived samples from the same individuals were compared (*within-individual, between tissue*) and from different members of the same MZ pair (*between co-twins, within tissue* and *between co-twins, between tissue*). The consistency between and within DZ pairs was also examined. To assess the consistency of CN inference over time, longitudinal samples of the same tissue type (*within-individual, between time points*) and between MZ co-twins were compared (*between co-twins, within time* and *between co-twins, between time points*).

For a concordance metric, we applied the squared Pearson correlation (R^2), a commonly used summary for imputation quality. This statistic has the advantage of being relatively sensitive to important differences in genotype, regardless of minor allele frequency (Huang et al., 2009), especially compared with a simple concordance of values. Studies of CN inference often assume particular criteria for designating putative CN segments as 'real' for analysis (e.g., two segments may require 50% overlap of one of the segments; some regions of duplications of differing copy number, such as 3 and 4, may be considered identical). These criteria are arbitrary in nature. The choice to evaluate consistency between two samples with R^2 across calls at all probes, and across copy number, effectively integrates

over choices for rules of sufficient segmentation overlap as well as CN-based categorization of duplication and deletion event identity. The result of this procedure should render our results more generalizable, and the ordering of accuracies more robust to assumptions.

Materials and Methods

Sample

This study consisted of 1,472 samples from 710 individuals collected from the NTR (Boomsma et al., 2006). Subjects had their blood drawn at ages 4–31 years (see Table 1 for more details). The longitudinal samples from the same individual were collected at 1–12-year intervals. Zygosity was confirmed by concordance through interviews and the genome-wide Affymetrix 6.0 data, using an identical-by-descent (IBD) measure from PLINK and applying 1.8 as a cut-off for zygosity determination.

Ethics

The study was approved by the Central Ethics Committee on Research involving human subjects of the VU University Medical Centre, Amsterdam, an institutional review board certified by the U.S. Office of Human Research Protections (IRB number IRB-2991 under Federal-wide Assurance-3703; IRB/institute codes, NTR 03-180).

DNA Collection

Buccal

Participants provided buccal swabs for DNA extraction, collected at ages 4–52 (see Table 1). Methods for buccal swab collection and genomic DNA extraction have been described previously (Meulenbelt et al., 1995; Willemsen et al., 2010). In brief, 16 cotton mouth swabs were individually rubbed against the inside of the cheek by the participants and placed in four separate 15 mL conical tubes (four swabs in each tube) containing 0.5 mL STE buffer (100 mM Sodium Chloride, 10 mM Tris Hydrochloride [pH 8.0], and 10 mM Ethylenediaminetetraacetic acid) with proteinase K (0.1 mg/mL) and Sodium Dodecyl Sulfate (SDS) (0.5%) per swab. Individuals were asked to refrain from eating or drinking 1 hour prior to sampling. High molecular weight genomic DNA was extracted from the swabs using a high salt (KAc) precipitation followed by a standard chloroform/isoamyl alcohol (24:1) extraction. The DNA samples were quantified using absorbance at 260 nm with a Nanodrop ND-1000 (Nanodrop Technologies, Wilmington, DE, USA).

Whole Blood

Blood samples were derived from two projects that took place around 1995 (van Dijk et al., 1996) and around 1997–98 (Estourgie-van Burk et al., 2010). DNA samples from buffy coats were extracted from the first study and DNA samples from EDTA tubes in the second project (see Boomsma et al., 2008; Willemsen et al., 2010, for description of DNA isolation). High molecular weight genomic DNA from whole blood and buffy coats was extracted using the QIAamp DNA Blood Maxi kit (QIAGEN; Dusseldorf, Germany). The DNA concentrations were normalized using absorbance at 260 nm.

Genotyping and Quality Control

SNP and CNV data were generated from each DNA sample using high-density SNP 6.0 micro-arrays (Affymetrix; Santa Clara, CA, USA). This array represents 1.8 million markers evenly distributed across the whole genome, including more than 906,600 markers to detect SNPs and another 946,000 non-polymorphic probes for the detection of CNVs. Sample preparation and micro-array processing were performed according to standard protocols. Per sample data quality metrics were gauged using a single analysis suite, Genotyping Console v4.0 (Affymetrix; Santa Clara, CA, USA), which includes a high-resolution reference map and a CN polymorphism calling algorithm. Samples were plated according to our own family-based plating approach. All twins, siblings, and parent samples were included on same plate without exception. Blood and buccal samples from the same individual were also on the same plate. Several internal controls were included on each plate to assist in measuring data quality. Each plate of genotyped samples included five internal controls. The Affymetrix control (Ref103) was included on each plate, which represents genomic DNA from a male sample generated from a stable cell line. In addition, on each plate, a well-characterized female and male buccal-derived genomic DNA sample was included. On alternating plates, a blood-derived genomic DNA from either the same female or male sample used as the buccal-derived genomic DNA control was included. Lastly, on alternating plates, a within-plate or between-plate control was included as well.

Affymetrix Genotyping Console (GTC) 4.0 was used to calculate Contrast Quality Control (CQC) and Median of the Absolute values of all Pairwise Differences (MAPD) metrics for a measure of per sample quality control (QC). CQC is a metric for how well the allele intensities separate into clusters, with lower CQC values indicating a higher difficulty for the algorithm in distinguishing homozygotic genotypes from heterozygotic genotypes. MAPD is an estimate of variability or standard deviation, where increased variability decreases the quality of CN calls. Samples with a CQC < 0.4 or an MAPD > 0.35 were excluded for analyses of genotypes. The data set is considered problematic if more than 10% of the samples do not pass the CQC cut-off of 0.4 or when the mean CQC is smaller than 1.70, which was not observed in our plates.

Genotype Calling

Genotypes from the Affymetrix 6.0 CEL files were generated using Canary of Birdsuite v.2 (Korn et al., 2008). We also used the GTC v.4 software, but found only negligible

differences between calls from the same samples (data not shown). Genotypes were called at each SNP (marginally).

Copy Number Inference

We applied the following two methods for CN inference: Birdsuite v.2 and PennCNV (Wang et al., 2007). For tissue-specific and longitudinal comparisons, we used results from Birdseed, excluding calls for which the 'logarithm (base 10) of odds (LOD) scores' were less than 10. For R^2 calculations, we converted all CN calls to probewise sample-specific CN designations. To facilitate comparisons between samples, where a program failed to make a non-copy-neutral call, a value of '2' (copy neutral) was inserted. The 'log R measure' is the $\log_2$ ratio of the probe intensity to the median probe intensity across samples on the same plate.

Statistical Quality Control

We applied statistical QC filters to the 1,471 samples in this study based on criteria established using 4,801 samples. We first calculated per-sample missing data rates. Next, using all samples from individuals in the same family, we performed 'identity-by-descent' and identity-by-state calculations (using PLINK; Purcell et al., 2007) on all pairwise combinations of samples. Then, excluding further individuals, we suspected of being involved in any type of sample mix up (based on the genome-wide data), we calculated a series of per-SNP QC steps to derive a final data set by requiring a marker to pass *all* of the following criteria: (1) genotype no-call rate <5%; (2) Hardy–Weinberg equilibrium p value $> 10^{-5}$; (3) total Mendelian inconsistencies < 27 (across 291 families); (4) genotype discordance between sample duplicates <5%; and (5) minor allele frequency (MAF) $>$ 0.00141 (corresponding to at least two minor alleles among unrelated individuals). After applying these filters, we retained 827,983 (of 882,242 or approximately 94%) SNPs.

Comparisons of CNV Inference and SNP Genotyping Across Samples

We calculated R^2, the square of the Pearson correlation coefficient, between CN and genotype calls between two samples. To compare two groups of samples, each containing individual genotype and CN calls, we reported the median R^2 value among all pairs in each contrast.

Statistical Analysis

For several comparisons, we stated statistical significance between two distributions of R^2 values. These were conducted with non-parametric Wilcoxon rank-sum tests, using two-sided alternative hypotheses. Where appropriate, a paired test was applied (sign-rank test); and when possible (no ties), an exact test of significance was obtained. When evaluating genotype concordance for 'Raw' (unfiltered) genotypes, comparing blood–blood concordances against two other groups, we obtained a permutation-based p value from 10,000 shuffles of the pairwise R^2 values among

tissue types within a unit (pair of individuals); in all permutations the difference in R^2 between the largest and next-largest median values was less than 0.004, the observed difference. To test for differences in allelic intensities (i.e., in Figure 1), we applied a t test.

Results

Our results show CN calls from blood and buccal tissues from the same individual to be highly concordant. Table 2 contains results on a complete set of comparisons between tissues (blood and buccal) across probe types. As it is helpful to make overall comparisons on the same sets of individuals, we focus on MZ pairs. This generates both blood and buccal replicates from a priori genetically identical units, facilitating across-tissue comparisons where it is desirable to compare the same sets of individuals. For example, the blood–buccal consistency within MZ pairs is on the order of that for blood–blood and buccal–buccal contrasts, especially for deletions. However, there is a slight decrease (0.03–0.04 absolute difference in R^2) in between-tissue concordance (blood–buccal is lower than buccal–buccal and blood–blood for 'All CNVs'; $p = .005$).

We test the assumption of genetic identity of MZ co-twins separately. The 'MZ within-individual' results in Table 2 show the within-individual variation to be indistinguishable from comparisons between MZ co-twins, suggesting that this analysis is suitable for the purposes of testing consistency of CN inference from blood and buccal sources. In fact, the observed difference is in the direction of less concordance within individuals (88–89% within and 89–90% between). Of course, we do expect some CN differences in CNVs between MZ co-twins, due to post-twinning, somatic changes in copy number. (We have investigated the effects of such phenomena on behavioral changes in Ehli et al., 2012.) The within-person concordances of the DZ pairs are approximately the same as for the MZ pairs. As expected, the between DZ-twin comparisons, including those of the same tissue, are lower, due to the fact that these twins should have the genetic similarity of full siblings.

For comparisons within the same individual or between MZ co-twins, there is a slight decrease in the concordance for duplications when compared with deletions, indicating the duplications are either more difficult to reliably infer or more likely to demonstrate somatic mutation. We also note a difference in the direction of effect for the probe type between the CN classes. Specifically, CN probes show higher consistency at deletions (by ∼2%, absolute) and SNP probes show higher consistency at duplications (by ∼8%), possibly indicating that the bivariate allelic signal at SNP arrays is more informative for correlating or calling duplications than deletions.

We note that the ranks of probe classes (SNP vs. CN), based on median R^2 values, change based on the specific combinations of CN variation (deletion vs. duplication)
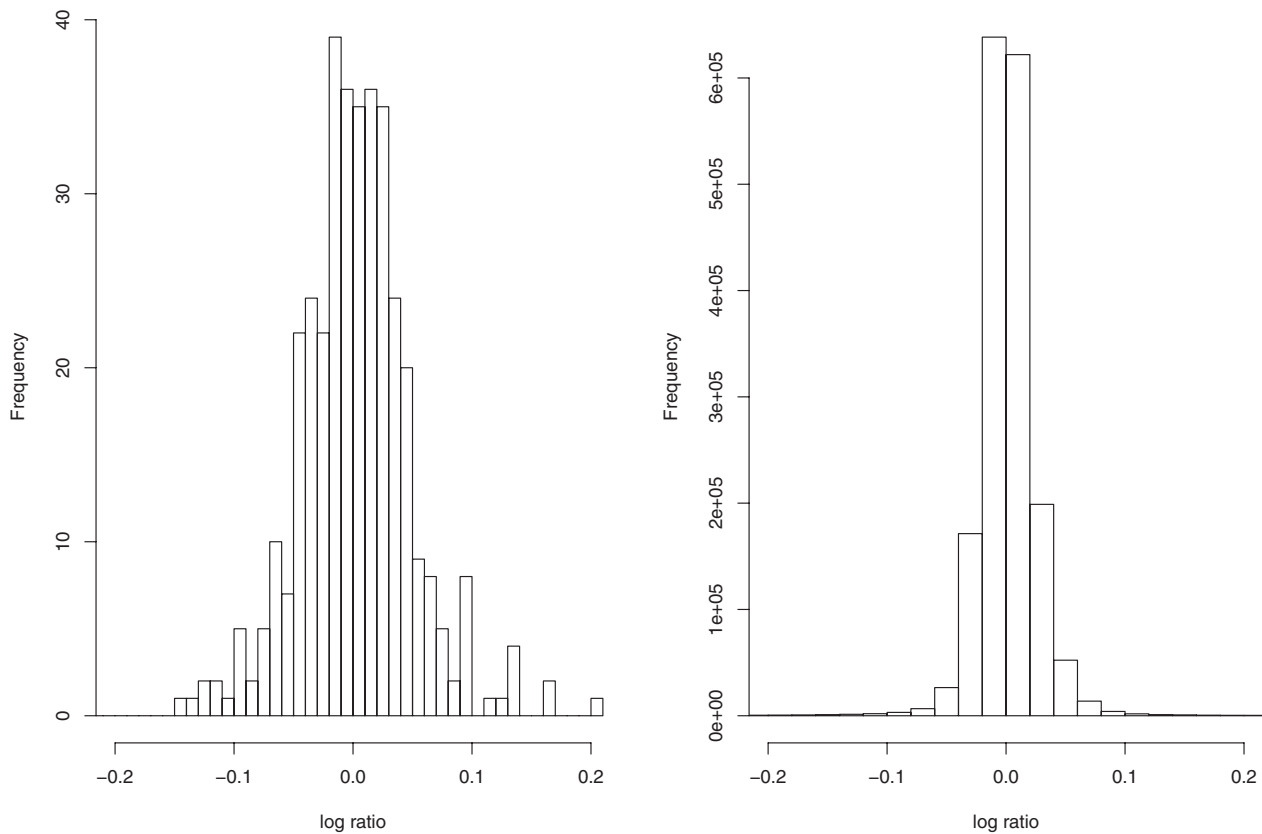
FIGURE 1

Average differences in $\log_2$ intensities from blood and buccal DNA sources. Differences are computed as the $\log_2$ ratio of total intensity from blood samples to intensity from buccal. Left: differences averaged over probes (372 values in histogram). Right: differences averaged over individuals (1.7M values plotted). The mean shift is 0.0012 ($p < .0001$).

and the zygosity of individual comparisons. For example, the between DZ co-twin contrasts show greater consistency for CN probes than for SNP probes (∼56% vs. ∼46%). This can be explained by the correlation induced by population variation at regions of the genome where none of the chromosomes are IBD. Indeed, unrelated individuals exhibit increased variation at the CN probes, where the loci have been identified based on population studies of CN polymorphism. The rare CNVs picked up with SNP probes will tend to be more rare and thus are less likely to be shared 'by chance' between unrelated individuals (or regions where zero chromosomes are IBD in full siblings or DZ twins; see Table S1 of the Supplementary Material for additional evidence of this effect). This phenomenon does not occur for comparisons of samples from the same individual or from MZ co-twins, as for these sample comparisons, similarities and differences can be ascribed to one source of variation only (i.e., measurement error).

Importantly for GWA studies, genotype accuracies are generally high between the two tissue types, with R^2 values above 99% (Table 2). There appeared a slightly higher concordance for blood–blood comparisons than for buccal–buccal or buccal–blood comparisons (0.995 vs. 0.990 and 0.991).

CN events were also determined by making a non-neutral CN call if and only if the same call (i.e., deletion or duplication) was called by *two* algorithms, namely Birdsuite and PennCNV (Wang et al., 2007). This resulted in an increased concordance of CN calls of about 7% in absolute terms (Table S3). Our conclusion of no consistent difference in CN inference from different tissue sources remained.

Because these CN calling algorithms rely on micro-array intensities, individual probe intensities were also evaluated directly. We used the 'log R measure', which is normalized by plate. Figure 1 shows distributions of differences between blood and buccal intensities. No evidence of systematic differences between tissue types in terms of raw intensities was seen, consistent with our conclusions of robustness of CNV inference. These conclusions hold if we first excluded inferred deletions, because intensity ratios in these segments may be more variable.

In 388 unique individuals in the NTR study who had DNA collected at multiple time points, we assessed stability across time. At least 3 years separated 130 sample pairs (from same individual), and 122 sample pairs were separated by at an interval of at least 7 years. Twenty-three sample pairs were separated by 12 years, the largest interval (see Table 3). We regressed CNV concordance by time interval and

TABLE 2

Concordance of CNVs and Genotypes Across Blood and Buccal DNA Sources

Contrast	n	All CNVs	Deletions SNP	Deletions CN	Duplications SNP	Duplications CN	Genotypes Raw	Genotypes Filtered	n
Within individual									
Buccal − Blood	372	0.890	0.929	0.949	0.901	0.810	0.991	0.995	331
MZ (within individual)									
Blood − Buccal (twin 1)	43	0.876	0.940	0.955	0.900	0.807	0.992	0.995	40
Blood − Buccal (twin 2)	43	0.886	0.928	0.953	0.903	0.816	0.993	0.996	39
MZ (co-twin associations)									
Buccal (1) − Blood (2)	43	0.898	0.933	0.963	0.902	0.809	0.991	0.996	40
Buccal (2) − Blood (1)	43	0.889	0.928	0.952	0.904	0.821	0.991	0.995	39
Buccal (1) − Buccal (2)	43	0.916	0.931	0.966	0.923	0.855	0.990	0.994	39
Blood (1) − Blood (2)	43	0.920	0.934	0.960	0.917	0.861	0.995	0.997	41
DZ (within individual)									
Blood − Buccal (twin 1)	75	0.897	0.928	0.947	0.906	0.835	0.990	0.994	70
Blood − Buccal (twin 2)	75	0.894	0.930	0.947	0.901	0.819	0.987	0.993	64
DZ (co-twin associations)									
Buccal (1) − Blood (2)	75	0.595	0.449	0.558	0.458	0.547	0.654	0.636	70
Buccal (2) − Blood (1)	75	0.614	0.438	0.553	0.478	0.584	0.654	0.640	63
Buccal (1) − Buccal (2)	75	0.616	0.437	0.553	0.463	0.591	0.650	0.638	61
Blood (1) − Blood (2)	75	0.609	0.458	0.569	0.495	0.567	0.662	0.641	73

Note: Sample R^2 is calculated for all contrasts, either from the same individual or twin sibship, and is reported as the *median* among comparison groups. Siblings within a twin pair are arbitrarily labeled '1' and '2'. 'All CNVs' summarizes concordance across all CN states (0–6 for Birdsuite). For the breakdown of CNV probe types (*SNP* for 'S-labeled' probes and *CN* for 'C-labeled'), we grouped all duplication (respectively, deletion) states into a single category of duplication (deletion). All blood and buccal samples from the same individuals or twin pair were plated together. When evaluating deletions, the presence of duplications was ignored (effectively treated as copy-neutral); likewise, deletions were ignored when evaluating duplications. Genotype concordance was calculated for all SNPs ('Raw') using all available samples, as well as for a set of post-QC SNPs ('Filtered') on a subset of individuals that passed preliminary per-sample QC metrics, that is, CQC (<0.4) and MAPD (>0.35) criteria. The number of subjects in comparisons is the same within rows, except for the 'Filtered' genotypes, which varies by contrast due to some subjects being excluded. Sample sizes for these contrasts are listed in the last column.

considered dichotomous groupings of time (e.g., at least 7 years vs. less than 7 years). No greater differences in concordance of estimated CNVs among individuals whose two samples were obtained at larger time intervals were found. (Plate effects were statistically adjusted for, as well.)

In addition, we examined concordance in MZ co-twins of different ages, that is, at approximately 9 or 18 years of age, across two time points. Twice the age of the twins represents the number of cumulative years for somatic variation to accumulate since the zygote was formed, and the two time points provide an additional contrast. We combined all possible contrasts (different ages, between twins; different time points, same individual; across both individuals and time) and regressed R^2 against time or years since zygotic stage, calculating coefficients ('slopes'). The distributions of these slopes center at zero (0.0007; Table 3), consistent with no systematic temporal effect.

To identify a possible 'positive control' where buccal- and blood-derived DNA would be expected to exhibit differences, we analyzed 779 probes, from regions on chromosomes 7 and 14, coding T-cell receptors (α, β, γ, δ) (Table S2 in the Supplementary Material). We compared probe intensities, CNV calls, and genotypes between MZ co-twins using blood and buccal DNA sources. As variable regions of T-cell receptors are known to rearrange in lymphocytes, a source for blood-derived DNA, we examined whether differences existed in inferred CN between blood samples (as they are rearranging due to immune response), but did not exist

TABLE 3

Concordance of CNVs for Longitudinal Samples

Interval between sampling (in years)	n	Median R^2 (SD)
1	2	0.947 (0.047)
2	257	0.918 (0.166)
3	8	0.844 (0.194)
7	65	0.927 (0.134)
8	14	0.948 (0.040)
10	1	0.892 (n/a)
11	19	0.868 (0.319)
12	23	0.903 (0.116)

MZ quads (2 co-twins, 2 time points: t_1, t_2)	(44)	Time contrast
Within individual; t_1 vs. t_2	88	$t_2 - t_1$
Within time t_1, twin 1 vs. twin 2	44	$2t_1$
Within time t_2, twin 1 vs. twin 2	44	$2t_2$
Both: twin 1/t_1, twin 2/t_2; twin 1/t_1, twin 2 / t_2	88	$2(t_1 + t_2)$

		Mean slope
All possible contrasts	264	0.0007

Note: Top: R^2 values were calculated for each pair of samples from the same individual and median values (and standard deviations) reported by time interval; for example, 257 individuals had samples taken at a 2-year interval and so on. Bottom: Forty-four MZ twin pairs had samples measured at two time points. There are six possible two-way comparisons among the sample pairs; each comparison generates a total time contrast (either from multiple samples of the same individual or from two MZ co-twins separated by their combined ages). Each contrast generated a R^2 value of concordance among CNVs and a time contrast. The mean slope is an average among the slopes obtained from individual linear regressions of R^2 on time contrast.

between buccal samples (as they are genetically identical by germline DNA). To assess this, we extracted normalized intensities ('log R measure'; from PennCNV), as well

TABLE 4

Genomic Consistency at T-Cell Receptors Across Tissue Types

Tissue contrast	log R measure	CN estimates	Genotypes
Blood–blood	0.336 (0.301)	0.954 (1.0)	0.990 (0.995)
Buccal–buccal	0.282 (0.218)	1.0 (1.0)	0.986 (0.991)
p*	.027	.004	.12

Note: We assessed genomic features at T-cell receptor loci and summarized the differences between MZ co-twins using DNA derived from blood and buccal. Average (median) R^2 values were computed across 43 MZ pairs (individual R^2 values calculated within MZ sibships, across probes). *P values were obtained from a Wilcoxon rank-sum test for differences in the distributions of R^2.

as CN and genotype calls (from Birdsuite), and then examined the differences between co-twins, summarized by tissue source for DNA. Table 4 contains average differences between MZ pairs for blood- and buccal-derived DNA. CN variation between co-twins was consistent with the hypothesis of greater similarity between DNA derived from buccal tissue than from blood ($p = .004$). Analyzing a subset of the markers within highly variable 'VDJ' regions, we obtained a slightly more significant result ($p = .0005$). We do observe marginal significance as well for the log R measure, but in the opposite direction (buccal-derived samples are less consistent); however, these average correlations are much lower and may reflect considerable noise.

A shift in CQC was seen toward smaller values for buccal samples (mean difference 0.46; $p < .0001$). As for these samples blood and buccal DNA were collected at the same time, we ruled out an 'age effect' for this phenomenon. Rather, this effect may be hypothesized as due to interference of human DNA hybridization from bacterial DNA. This shift did not appear to explain the subtle but statistically significant difference in R^2 values between same and different tissue types, because when comparisons were stratified by CQC values, this effect remained.

Accuracy of CN or genotype calls was not directly compared. Rather, we focused on *concordance*, or consistency, of calls, evaluating variance in estimates, instead of bias. However, due to study design (making comparisons across MZ twins *and* DNA sources) we can rule out situations that would mask differential accuracies. The focus of our study was on the appropriateness of the use of different tissue types and samples from twins for GWA studies, rather than a thorough attempt to catalog all evidence for somatic variation. Indeed, such phenomena have been characterized recently, with somatic variation detected between MZ twins (Forsberg et al., 2012) and within individuals (Laurie et al., 2012) at older ages, but a minimal or negligible effect in younger individuals. Our data set consists of samples from younger individuals. For example, the comparisons across MZ twins involved samples at a median age of 18 years, and individuals of the longitudinal comparisons had a median age at sampling of 10 years.

Finally, we examined consistency for SNP genotype and CNV inference across tissue types and twin pairs in a com-

plementary manner. We calculated R^2 at each probe by examining pairs of samples (e.g., from the same individual, across tissue types, or between MZ twins). In Table S5, we report various quantiles of these values among markers at which there existed some variation (so that a correlation could be calculated). For all comparisons involving MZ twins, the first quartile (more than three-quarters of sites) R^2 was above 90%. One subset of comparison (372 pairs of blood–buccal samples from the same individuals) showed a first quartile R^2 of 85%.

Discussion

The results across the tissue and longitudinal replicates demonstrated that buccal-derived DNA samples are suitable for GWA studies of SNP and CNV, ruling out any substantial effect on CN inference due to bacterial contamination from buccal sources. Current and emerging technologies, such as massively parallel DNA sequencing, may provide sufficient sensitivity to detect subtle phenomena, and studies that may use these technologies will require appropriate planning based on relevant experiments that explore this issue. In light of our data and results, it is acceptable, and indeed potentially cost effective, to conduct a micro-array-based GWA study of genotypes and CN using buccal-derived DNA for efforts to map genetic variants predisposing to disease and behaviors. Buccal swabs can be easier to collect from young (or very old) participants for whom requiring a blood draw for participation could be a barrier, and for individuals who have had bone marrow transplants or are undergoing chemotherapy and hence have low white blood cell counts. However, as there exists a subtle but consistent differential in concordance across tissue types, caution is advised when combining sources into a single analysis. Finally, our study confirms that collections of blood-derived DNA from large twin registries offer a valuable resource for mapping and characterizing complex diseases, given their unique contribution to partitioning environmental and genetic effects and the stability of genomic features across samples and time.

Acknowledgments

Paul Scheet et al.

References

Bartels, M., van Beijsterveldt, C. E., Derks, E. M., Stroet, T. M., Polderman, T. J., Hudziak, J. J., & Boomsma, D. I. (2007). Young Netherlands Twin Register (Y-NTR): A longitudinal multiple informant study of problem behavior. *Twin Research and Human Genetics, 10*, 3–11.

Boomsma, D. I., de Geus, E. J., Vink, J. M., Stubbe, J. H., Distel, M. A., Hottenga, J. J., Posthuma, D., van Beijsterveldt, T. C., Hudziak, J. J., Bartels, M., & Willemsen, G. (2006). Netherlands Twin Register: From twins to twin families. *Twin Research and Human Genetics, 9*, 849–857.

Boomsma, D. I., Willemsen, G., Sullivan, P. F., Heutink, P., Meijer, P., Sondervan, D., Kluft, C., Smit, G., Nolen, W. A., Zitman, F. G., Smit, J. H., Hoogendijk, W. J., van Dyck, R., de Geus, E. J., & Penninx, B. W. (2008). Genome-wide association of major depression: Description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *European Journal of Human Genetics, 16*, 335–342.

Dellinger, A. E., Saw, S. M., Goh, L. K., Seielstad, M., Young, T. L., & Li, Y. J. (2010). Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. *Nucleic Acids Research, 38*, e105.

Dlugos, D. J., Scattergood, T. M., Ferraro, T. N., Berrettinni, W. H., & Buono, R. J. (2005). Recruitment rates and fear of phlebotomy in pediatric patients in a genetic study of epilepsy. *Epilepsy & Behavior, 6*, 444–446.

Ehli, E. A., Abdellaoui, A., Hu, Y., Hottenga, J. J., Kattenberg, M., van Beijsterveldt, T., Bartels, M., Althoff, R. R., Xiao, X., Scheet, P., de Geus, E. J., Hudziak, J. J., Boomsma, D. I., & Davies, G. E. (2012). *De novo* and inherited CNVs in MZ twin pairs selected for discordance and concordance on Attention Problems [published online ahead of print]. *European Journal of Human Genetics, 20*, 1037–1043.

Erlich, Y. (2011). Blood ties: Chimerism can mask twin discordance in high-throughput sequencing. *Twin Research and Human Genetics, 14*, 137–143.

Estourgie-van Burk, G. F., Bartels, M., Boomsma, D. I., & Delemarre-van de Waal, H. A. (2010). Body size of twins compared with siblings and the general population: From birth to late adolescence. *The Journal of Pediatrics, 156*(4), 586–591.

Forsberg, L. A., Rasi, C., Razzaghian, H. R., Pakalapati, G., Waite, L., Thilbeault, K. S., Ronowicz, A., Wineinger, N. E., Tiwari, H. K., Boomsma, D., Westerman, M. P., Harris, J. R., Lyle, R., Essand, M., Eriksson, F, Assimes, T. L., Iribarren, C., Strachan, E., O'Hanlon, T. P., Rider, L. G., Miller, F. W., Giedraitis, V., Lannfelt, L., Ingelsson, M., Piotrowski, A., Pedersen, N. L., Absher, D., & Dumanski, J. P. (2012). Age-related somatic structural changes in the nuclear genome of human blood cells. *American Journal of Human Genetics, 90*, 217–288.

Huang, L., Li, Y., Singleton, A. B., Hardy, J. A., Abecasis, G., Rosenberg, N. A., & Scheet, P. (2009). Genotype-imputation accuracy across worldwide human populations. *The American Journal of Human Genetics, 84*, 235–250.

Korn, J. M., Kuruvilla, F. G., McCarroll, S. A., Wysoker, A., Nemesh, J., Cawley, S., Hubbell, E., Veitch, J., Collins, P. J., Darvishi, K., Lee, C., Nizzari, M. M., Gabriel, S. B., Purcell, S., Daly, M. J., & Altshuler, D. (2008). Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nature Genetics, 40*, 1253–1260.

Laurie, C. C., Laurie, C. A., Rice, K., Doheny, K. F., Zelnick, L. R., McHugh, C. P., Ling, H., Hetrick, K. N., Pugh, E. W., Amos, C., Wei, Q., Wang, L. E., Lee, J. E., Barnes, K. C., Hansel, N. N., Mathias, R., Daley, D., Beaty, T. H., Scott, A. F., Ruczinski, I., Scharpf, R. B., Bierut, L. J., Hartz, S. M., Landi, M. T., Freedman, N. D., Goldin, L. R., Ginsburg, D., Li, J., Desch, K. C., Strom, S. S., Blot, W. J., Signorello, L. B., Ingles, S. A., Chanock, S. J., Berndt, S. I., Le Marchand, L., Henderson, B. E., Monroe, K. R., Heit, J. A., de Andrade, M., Armasu, S. M., Regnier, C., Lowe, W. L., Hayes, M. G., Marazita, M. L., Feingold, E., Murray, J. C., Melbye, M., Feenstra, B., Kang, J. H., Wiggs, J. L., Jarvik, G. P., McDavid, A. N., Seshan, V. E., Mirel, D. B., Crenshaw, A., Sharopova, N., Wise, A., Shen, J., Crosslin, D. R., Levine, D. M., Zheng, X., Udren, J. I., Bennett, S., Nelson, S. C., Gogarten, S. M., Conomos, M. P., Heagerty, P., Manolio, T., Pasquale, L. R., Haiman, C. A., Caporaso, N., & Weir, B. S. (2012). Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nature Genetics, 44*, 642–650.

Meulenbelt, I., Droog, S., Trommelen, G., Boomsma, D., & Slagboom, P. (1995). High-yield noninvasive human genomic DNA isolation method for genetic studies in geographically dispersed families and populations. *American Journal of Human Genetics, 57*, 1252–1254.

Min, J. L., Lakenberg, N., Bakker-Verweij, M., Suchiman, E., Boomsma, D. I., Slagboom, P. E., & Meulenbelt, I. (2006). High microsatellite and SNP genotyping success rates established in a large number of genomic DNA samples extracted from mouth swabs and genotypes. *Twin Research and Human Genetics, 9*, 501–506

Pinto, D., Darvishi, K., Shi, X., Rajan, D., Rigler, D., Fitzgerald, T., Lionel, A. C., Thiruvahindrapuram, B., Macdonald, J. R., Mills, R., Prasad, A., Noonan, K., Gribble, S., Prigmore, E., Donahoe, P. K., Smith, R. S., Park, J. H., Hurles, M. E., Carter, N. P., Lee, C., Scherer, S. W., & Feuk, L. (2011). Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nature Biotechnology, 29*, 512–520.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics, 81*, 559–575.

Rincon, G., Tengvall, K., Belanger, J. M., Lagoutte, L., Medrano, J. F., André, C., Thomas, A., Lawley, C. T., Hansen, M. S., Lindblad-Toh, K., & Oberbauer, A. M. (2011). Comparison of buccal and blood-derived canine DNA, either native or whole genome amplified,

for array-based genome-wide association studies. *BMC Research Notes*, *4*, 226.

Talens, R. P., Boomsma, D. I., Tobi, E. W., Kremer, D., Jukema, J. W., Willemsen, G., Putter, H., Slagboom, P. E., & Heijmans, B. T. (2010). Variation, patterns, and temporal stability of DNA methylation: Considerations for epigenetic epidemiology. *The FASEB Journal*, *24*, 3135.

van Dijk, B. A., Boomsma, D. I., & de Man, A. J. M. (1996). Blood group chimerism in human multiple births is not rare. *American Journal of Medical Genetics*, *61*, 264–268.

Willemsen, G., de Geus, E. J., Bartels, M., van Beijsterveldt, C. E., Brooks, A. I., Estourgie-van Burk, G. F., Fugman, D. A., Hoekstra, C., Hottenga, J. J., Kluft, K., Meijer, P., Mont-gomery, G. W., Rizzu, P., Sondervan, D., Smit, A. B., Spijker, S., Suchiman, H. E., Tischfield, J. A., Lehner, T., Slagboom, P. E., & Boomsma, D. I. (2010). The Netherlands Twin Register Biobank: A resource for genetic epidemiological studies. *Twin Research and Human Genetics*, *13*, 231–245.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F., Hakonarson, H., & Bucan, M. (2007). PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Research*, *17*, 1665–1674.

Woo, J. G., Sun, G., Haverbusch, M., Indugula, S., Martin, L. J., Broderick, J. P., Deka, R., & Woo, D. (2007). Quality assessment of buccal versus blood genomic DNA using the Affymetrix 500 K GeneChip. *BMC Genetics*, *8*, 79.