Twin Research and Human Genetics

Supplementary Material

Twins, tissue and time: An assessment of SNPs and CNVs

P Scheet[1,*], EA Ehli[2,3,*], X Xiao[1], CEM van Beijsterveldt[4], A Abdellaoui[4], RR Althoff[5], JJ Hottenga[4] , G Willemsen[4], KA Nelson[2], P Huizenga[2], Y Hu[2], CI Amos[6], M Bartels[4], MM Groen-Blokhuis[4], EJC de Geus[4], JJ Hudziak[5], GE Davies[2,3], DI Boomsma[4]

1 Department of Epidemiology, University of Texas M. D. Anderson Cancer Center, Houston, TX, USA

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

3 Department of Psychiatry, University of South Dakota, Sanford School of Medicine, Division of Basic Biomedical Sciences, Sioux Falls, SD, USA

4 Department of Biological Psychology, VU University, Amsterdam, The Netherlands

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

6 Department of Genetics, University of Texas M. D. Anderson Cancer Center, Houston, TX, USA

Plate Effects. To rule out the possibility that results in Table 2 were inflated by our plating strategy in which samples from family members were plated together, we performed the following comparisons. We selected plates that contained (at least) 2 sets of MZ twin pairs. We then performed comparisons across families, calculating CN and genotype concordances among unrelated samples. We arbitrarily labeled the two MZ pairs from a plate as *A* and *B* (and members of a twin pair are labeled 1 and 2). Next, twin 1 from pair *A* was compared with twin 2 from pair *B*, and twin 2 from *A* with twin 1 from *B*, in a design analogous to results in Table 2 (main text). We did this with *A* and *B* from the same plate, as well as across plates, in both cases creating pairs from unrelated individuals. In all contrasts, since the samples are *a priori* unrelated according to records, any expected difference in concordances for within and between plate contrasts would be due to a "plate effect". Table S1 contains results of this experiment.

Table S1. *Median correlations of copy number and genotype calls between unrelated individuals*.

		All	**Deletions**		**Duplications**		**Genotypes**
Contrast (A1-B2 and A2-B1)	***n***	***CNVs***	**SNPs**	**CNPs**	**SNPs**	**CNPs**	
Same Plate							
Buccal – Blood (A1-B2 and A2-B1?	34	.390	.188	.313	.184	.384	.314
Buccal -- Buccal	17	.376	.194	.313	.185	.420	.313
Blood -- Blood	17	.353	.194	.316	.166	.333	.314
Different Plate							
Buccal -- Blood	34	.347	.176	.278	.222	.362	.317
Buccal -- Buccal	17	.363	.160	.292	.206	.398	.315
Blood -- Blood	17	.325	.185	.276	.231	.339	.318

We calculated R^2 for blood and buccal samples from unrelated individuals and report the *median* among comparison groups. As in Table 2, when evaluating one type of CN deviation, the presence of the other was ignored. For the buccal-blood comparisons, the median was averaged from the 2 possible groups -- buccal twin 1, blood twin 2; and buccal twin 2, blood twin1 – where 1 and 2 are arbitrary labels.

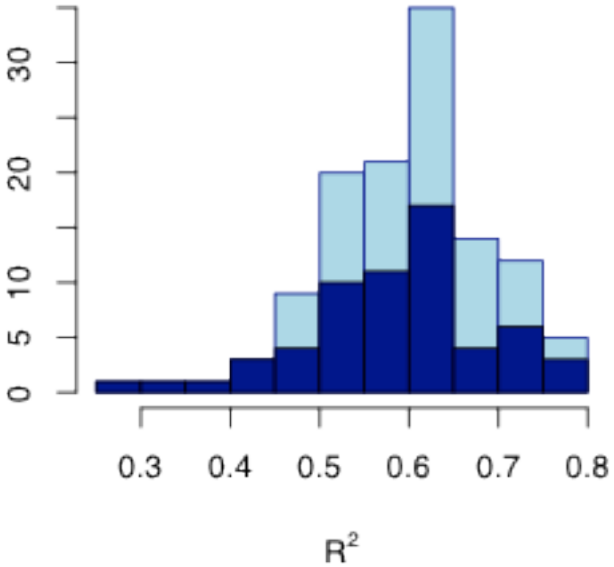
We see there is some evidence of a plate effect. Namely, the R^2 values for the sample comparisons on the same plate are higher than those from different plates, reflecting an *induced* correlation. However, the effect is modest, on the order of a 10% relative change. We also observe nonzero correlations of CN and genotype calls for unrelated individuals, along with a significantly higher correlation at CN probes, compared with SNP probes. For both SNP genotypes and CN, these effects are due to the correlation among sites induced by site-specific population SNP genotype allele and copy number variant frequencies. At sites where allele frequencies are higher, both unrelated individuals will be more likely to have the variant (SNP allele or CNV). At CNP probes, which were designed to measure "common" (or known) CNVs, there is increased correlation (over SNP CNV probes), since there is more leverage for this effect, ie. there are greater numbers of deviations in the same direction. The reason this differential (SNP vs. CN

probes) is not observed in Table 2 is that those samples are derived from the same individual or MZ sibship. In Table 2, population frequencies will not induce correlations since comparisons are made not on individuals from the same population but rather on samples from the same individual. Although this is a small comparison, the results are strikingly different from those in Table 2, yet sufficiently similar for within and between plates to eliminate plate effects as an overriding factor inducing similarities between co-twins and sample duplicates.

Examination of evidence for chimerism.

By examining SNP genotype concordance, we found no examples of the extreme situation where monozygosity would be indicated by analysis of DNA from blood but not buccal. In data from DZ twins genotyped for both tissue types, CN inference -- a potentially complementary and less stringent criterion -- was also evaluated to see if there existed a difference in concordance between tissues. We did not observe any shift in overall concordance towards differential similarity in co-twin comparisons based on DNA derived from different sources. Since we expect a modest number of twins to be affected by chimerism, we also examined samples in extreme quantiles rather than entire distributions. (Essentially, we are looking for a mixture of sources – those from chimeras and those not.) Here we did observe some differences among samples in the low quantile, but only for CNV concordance; we found no differences based on SNP genotypes.

Figure S1. *Pairwise CNV concordances between DZ cotwins using different tissues.*



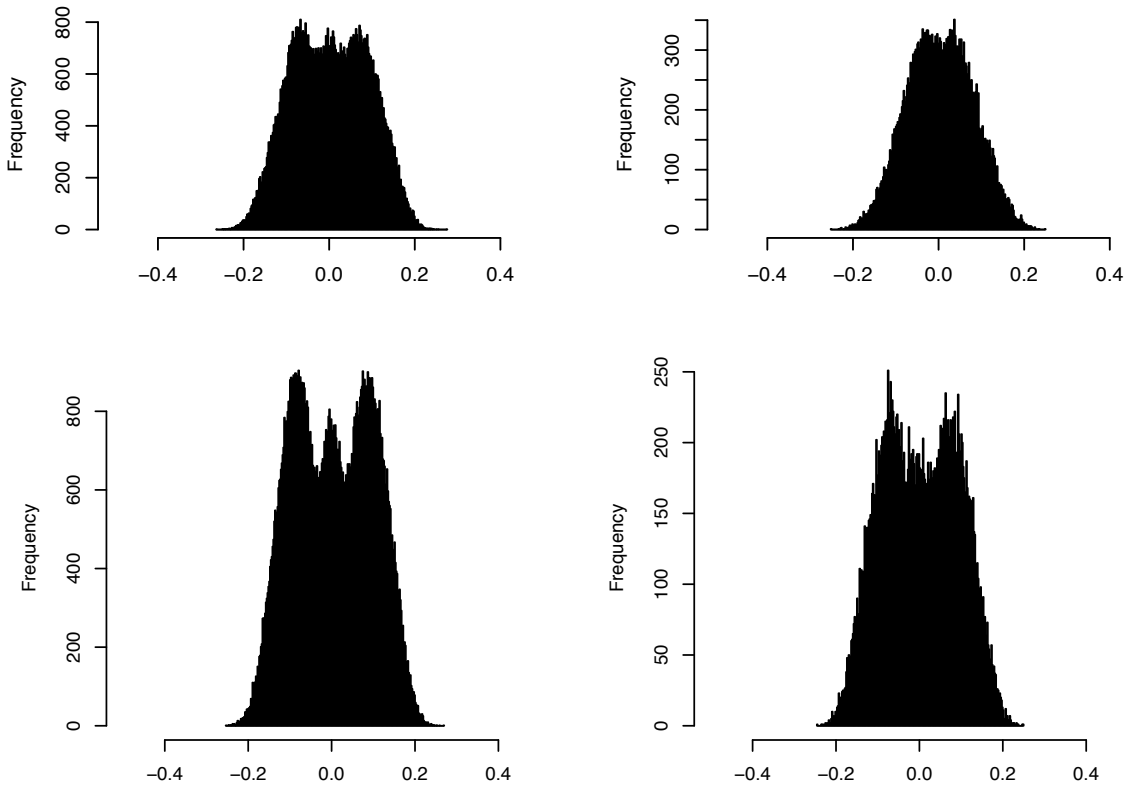
The combined area (bars) colored in *either* light or dark blue represent the distribution of all comparisons between DZ co-twins. The distributions for comparisons between co-twins using blood-derived DNA is indicated in dark blue shading.

The distribution of pairwise R^2 values for CNVs from DZ co-twins is displayed in Figure S1. The dark blue portion of the distribution highlights the comparisons

from blood, a subset of the total samples. We observe a shift towards smaller values for blood and a test of differences among the lowest ten R^2 values from the blood- and buccal-based co-twin comparisons is significant (P = .006). However, many of the samples involved in the comparisons that resulted in lower values were obtained for a specific study and thus had different storage conditions. If, when available, we substituted these samples with blood samples of the same individual collected from an older study, the difference did not remain significant. (That is, the effect may have been due to a confounding with data collection and storage conditions.)

Finally, we also note that by examining 115 families where we had measured buccal-derived DNA from DZ twins and at least one additional full sibling (not from the DZ pair), there was no greater concordance in CNVs (or SNP genotypes) for the DZ sibs compared with either co-twin and their non-twin full sibling.

Figure S2. *Intensity distributions across all SNP probes, stratified by similarity to co-twin for 2 individuals.*



Here we have plotted the "low level" data that forms the CQC (quality control) metric for 2 samples (1 sample per row). Well separated peaks indicate higher qualities of called genotypes (more resolution among genotype classes). LEFT: the distribution of intensities at SNP sites where the individual's genotype was identical to the co-twin's. RIGHT: the distribution at sites where they are different. (For each plot, random samples of probes were taken.)

Analysis of probes in T-cell receptors. The following genomic coordinates (based on hg18) were used to identify 779 probes from regions on chromosomes 7 and 14, coding T-cell receptors (α, β, γ, δ):

Table S2. *Locations of probes in regions of T-cell receptors*

Receptor	Chr	Pos. first probe	Pos. last probe
Gamma	7	38246749	38371496
Beta	7	141648712	142218889
Alpha/ delta	14	21164794	22090546

Multiple algorithms. We examined CNV calls made by the intersection of 2 algorithms: Birdsuite and PennCNV. Here we consider CN to be non-neutral if and only if it is called as a duplication by both algorithms, or if it is called to be a deletion by both algorithms (exact copy numbers within these classes are not distinguished). This procedure leads to higher overall concordances. Results are displayed in Table S2.

Table S3. *Median correlations of copy number at sites called by multiple algorithms*.

Contrast	***n***		**All CNVs**	By probe type	
				SNP	**CN**
ALL-Buccal-Blood	371	Birdsuite	.902	.915	.901
		Intersect.	**.963**	**.980**	**.959**
MZ only please specify which is within individual / between twins/					
Buccal-Blood	86	Birdsuite	.908	.909	.909
		Intersect.	**.971**	**.976**	**.967**
Buccal-Buccal	43	Birdsuite	.927	.927	.928
		Intersect.	**.982**	**.991**	**.982**
Blood-Blood	43	Birdsuite	.922	.920	.924
		Intersect.	**.983**	**.989**	**.983**

We calculated R^2 for blood and buccal samples from 371 individuals and report these for the 43 MZ twin sibships separately, as well. For reference we include results (not shown in the text), computed by Birdsuite only where we collapse copy number in the same manner (states 0 and 1 are collapsed, and states >2 are collapsed). For the buccal-blood comparisons here, we averaged the median from the 2 possible groups -- buccal twin 1, blood twin 2; and buccal twin 2, blood twin1 – where 1 and 2 are arbitrary labels.

In table S4 we tabulate CNVs by algorithm and copy number and give their size distributions.

Table S4. *Size and length distributions for copy number segments*.

N segments		*Copy number state (0-6)*							
		0	**1**	**2**	**3**	**4**	**5**	**6**	**>=3**
Birdsuite	Buccal	39	94	259	13	9	1	1	24
	Blood	39	93	295	14	9	1	1	24
PennCNV	Buccal	8	22	.	10	5	.	.	15 17
	Blood	8	22	.	11	6	.	.	
Total (genomic) length (in Mb)									
		0	**1**	**2**	**3**	**4**	**5**	**6**	**>=3**
Birdsuite	Buccal	24.48	3.45	2925	1.24	1.21	0.27	0.21	2.75
	Blood	24.83	3.38	2923	1.33	1.21	0.27	0.07	2.86
PennCNV	Buccal	0.17	1.23	.	1.17	0.14	.	.	
	Blood	0.18	1.09	.	1.52	0.18	.	.	
Typical size: (Total length) / (median N of segments) (in kb)									
		0	**1**	**2**	**3**	**4**	**5**	**6**	**>=3**
Birdsuite	Buccal	628	37	11293	96	135	270	207	115
	Blood	637	36	9910	95	135	274	67	119
PennCNV	Buccal	21	56	.	117	28	.	.	
	Blood	23	50	.	138	30	.	.	

For CN calls made by Birdsuite and PennCNV, we display here the median number segments, total genomic lengths, and total length divided by median number of segments (to yield a “typical” segment size) – for each CN state (0-6). These results are displayed for buccal and blood separately and computed on all 371 individuals.

Table S5. *Consistency as measured at each probe*. We calculated R^2 at each probe by examining pairs of samples (those in column 2 below). We then report various quantiles, ie the minimum (R2 P0), 5th (P5), 10th (P10), 25th (P25), 50th (P50, median), and 75th (P75) percentiles. We report results at all probes (SNP + CNP), as well as by each type individually. We also give results based on called genotypes (unfiltered and filtered). "N" denotes the sample size that went into the quantile calculations.

		CNV						
		N	**R2 P0**	**R2 P5**	**R2 P10**	**R2 P25**	**R2 P50**	**R2 P75**
All (372)	Buccal-Blood	62273	< .001	0.15	0.50	0.94	1	1
	Blood1-Buccal1	19106	0.001	0.48	0.67	1	1	1
MZ (43)	Blood2-Buccal2	19451	< .001	0.49	0.74	1	1	1
SNP+CNP	Buccal1-Blood2	19117	0.001	0.48	0.68	1	1	1
	Buccal2-Blood1	19520	< .001	0.43	0.73	1	1	1
	Buccal1-Buccal2	19425	< .001	0.49	0.73	1	1	1
	Blood1-Blood2	19911	< .001	0.42	0.75	1	1	1
	Blood1-Buccal1	29593	< .001	0.43	0.70	1	1	1
DZ (75)	Blood2-Buccal2	29576	< .001	0.34	0.67	1	1	1
SNP+CNP	Buccal1-Blood2	21911	< .001	0.00	0.05	0.20	0.49	1
	Buccal2-Blood1	22448	< .001	0.00	0.04	0.20	0.48	1
	Buccal1-Buccal2	22342	< .001	0.00	0.06	0.20	0.49	1
	Blood1-Blood2	22763	< .001	0.00	0.06	0.21	0.49	1
All (372)	Buccal-Blood	25934	< .001	0.40	0.63	1	1	1
	Blood1-Buccal1	5896	0.001	0.67	0.95	1	1	1
MZ	Blood2-Buccal2	6159	< .001	0.75	1	1	1	1
SNP	Buccal1-Blood2	5923	0.001	0.73	1	1	1	1
	Buccal2-Blood1	6070	0.001	0.73	0.95	1	1	1
	Buccal1-Buccal2	6099	0.001	0.69	1	1	1	1
	Blood1-Blood2	6216	< .001	0.73	1	1	1	1
	Blood1-Buccal1	10162	< .001	0.60	0.97	1	1	1
DZ	Blood2-Buccal2	9868	< .001	0.51	0.90	1	1	1
SNP	Buccal1-Blood2	6499	< .001	< .01	0.09	0.42	1	1
	Buccal2-Blood1	6618	< .001	< .01	0.09	0.41	1	1
	Buccal1-Buccal2	6646	< .001	< .01	0.09	0.39	1	1
	Blood1-Blood2	6799	< .001	< .01	0.11	0.43	1	1

All (372)	Buccal-Blood	36339	< .001	0.10	0.47	0.85	1	1
	Blood1-Buccal1	13210	0.001	0.27	0.56	0.96	1	1
MZ	Blood2-Buccal2	13292	0.001	0.27	0.67	1	1	1
CNP	Buccal1-Blood2	13194	0.001	0.29	0.60	1	1	1
	Buccal2-Blood1	13450	< .001	0.27	0.65	0.96	1	1
	Buccal1-Buccal2	13326	< .001	0.35	0.65	1	1	1
	Blood1-Blood2	13695	< .001	0.37	0.65	1	1	1
	Blood1-Buccal1	19431	< .001	0.31	0.55	0.98	1	1
DZ	Blood2-Buccal2	19708	< .001	0.27	0.54	0.95	1	1
CNP	Buccal1-Blood2	15412	< .001	< .01	0.04	0.18	0.37	1
	Buccal2-Blood1	15830	< .001	< .01	0.03	0.18	0.35	1
	Buccal1-Buccal2	15696	< .001	< .01	0.05	0.18	0.36	1
	Blood1-Blood2	15964	< .001	< .01	0.05	0.18	0.39	1

Genotypes

		N	R2 P0	R2 P5	R2 P10	R2 P25	R2 P50	R2 P75
All (372)	Buccal-Blood	808817	< .001	0.12	0.66	0.91	0.96	0.98
	Blood1-Buccal1	747651	< .001	0.73	0.88	1	1	1
MZ	Blood2-Buccal2	746215	< .001	0.74	0.89	1	1	1
SNP	Buccal1-Blood2	747272	< .001	0.73	0.88	1	1	1
	Buccal2-Blood1	746401	< .001	0.74	0.89	1	1	1
	Buccal1-Buccal2	755461	< .001	0.63	0.83	1	1	1
	Blood1-Blood2	744663	< .001	0.87	0.95	1	1	1
	Blood1-Buccal1	767232	< .001	0.56	0.81	0.95	1	1
DZ	Blood2-Buccal2	767522	< .001	0.49	0.76	0.93	1	1
SNP	Buccal1-Blood2	762127	< .001	0.04	0.09	0.16	0.24	0.33
	Buccal2-Blood1	763752	< .001	0.03	0.09	0.16	0.24	0.32
	Buccal1-Buccal2	782475	< .001	0.01	0.07	0.15	0.23	0.32
	Blood1-Blood2	755880	< .001	0.06	0.10	0.17	0.25	0.33

Genotypes-Filtered

		N	R2 P0	R2 P5	R2 P10	R2 P25	R2 P50	R2 P75
All (372)	Buccal-Blood	774908	< .001	0.25	0.75	0.92	0.97	0.98
	Blood1-Buccal1	721851	< .001	0.81	0.91	1	1	1
MZ	Blood2-Buccal2	720615	< .001	0.81	0.91	1	1	1
SNP	Buccal1-Blood2	721595	0.001	0.79	0.90	1	1	1
	Buccal2-Blood1	720714	0.001	0.82	0.92	1	1	1
	Buccal1-Buccal2	727510	0.001	0.73	0.87	1	1	1
	Blood1-Blood2	719547	< .001	0.90	0.95	1	1	1
	Blood1-Buccal1	739512	< .001	0.66	0.85	0.96	1	1
DZ	Blood2-Buccal2	739699	< .001	0.60	0.81	0.94	1	1
SNP	Buccal1-Blood2	734824	< .001	0.05	0.09	0.16	0.24	0.33
	Buccal2-Blood1	736240	< .001	0.04	0.09	0.16	0.24	0.32
	Buccal1-Buccal2	752254	< .001	0.02	0.08	0.15	0.24	0.32
	Blood1-Blood2	729406	< .001	0.06	0.10	0.17	0.25	0.33