



## **Supplemental Figure S2. Quantitation of D17Z1 variation using Southern blotting or PCR.**

A. Genomic DNA from diploid cell lines (U2OS or LCL-XX) and somatic cell hybrids containing single HSA17s (Z1\_2.6, Z1\_4.3) was digested to completion with EcoRI to release wild-type and variant HORs that were detected by hybridizing with a D17Z1 probe. Band sizes reflect the entire HOR, from wild-type 16-mers to the 9-mer that is representative of 13-mers containing the EcoRI SNP. The diploid cell lines studied here are unrelated to the diploid CEPH 1345 family cell lines. All blots were collected as white chemiluminescence bands on a black background; image was inverted to black bands on light background for presentation purposes only. See Methods for information on image quantification.

B. HOR variation was also measured using limited cycle PCR, followed by restriction digestion with EcoRI (Figure 2A) to reveal variant HORs of all sizes (16-, 15-,14-,13-,12-,11-mers and 16/13-mer SNP). This method was quicker and less laborious. Here, to verify that the method in Supplemental Figure S2A and the PCR-digestion approach were equivalent, D17Z1 variation was measured by Southern blotting (SB) and PCR in five cell lines. The data indicate that there is no statistically significant difference between variation calculated by PCR versus Southern blotting. The PCR assay was used for all subsequent variation measurements. Cell line names are included on the X axis. Four CEPH 1345 family members were included; the number after 1345 indicates which individual in the pedigree was analyzed (see Supplemental Figure S2C).

C. The presence or absence of the 13-mer/indel HOR is thought to correlate with epiallele formation, but quantitation of the different types of variation (wild-type versus 13-mer/indel versus SNP) in diploid individuals showed that both epiallele and non-epiallele individuals had similar amounts of 13-mer HORs. The plot shows that the proportion of arrays with two types

of variant HORs (13-mer and 16/13-mer SNP) was more highly associated with epialleles (asterisks above bar graphs). Because the HSA17 homologs cannot be distinguished in diploid samples, it is not possible to attribute specific amounts of each variant to a specific HSA17. Pedigree of the CEPH 1345 family is shown above the data plots. Circles represent females, squares denote males, and half shaded circles and squares represent D17Z1/D17Z1-B centromere heterozygotes (epiallele carriers).

D. Wild-type versus variant HORs from all Z1/Z1 CEN individuals in the CEPH 1345 family versus epiallele (Z1/Z1-B) individuals in the same family was averaged. The proportion of wild-type versus 13-mer EcoRI SNP HORs and the presence of epialleles, even in these diploid lines in which HSA17 homologs cannot be distinguished, is more statistically significant than the proportion of HORs with the 13-mer/indel only. Thus, the 13-mer SNP is highly associated with the HSA17 epiallele.

E. To test the correlation between polymorphic D17Z1 HORs and epialleles in additional diploid lines unrelated to the CEPH 1345 family, we measured the amount of D17Z1 variation in two unrelated diploid cell lines, one of which has epialleles (HT1080). There was a strong correlation between homogenous, invariant D17Z1 arrays and centromere assembly at D17Z1, whereas the epiallele line HT1080 showed 90% variation in D17Z1. The HT1080 line (fibrosarcoma) was studied at an early passage and was confirmed to be diploid.