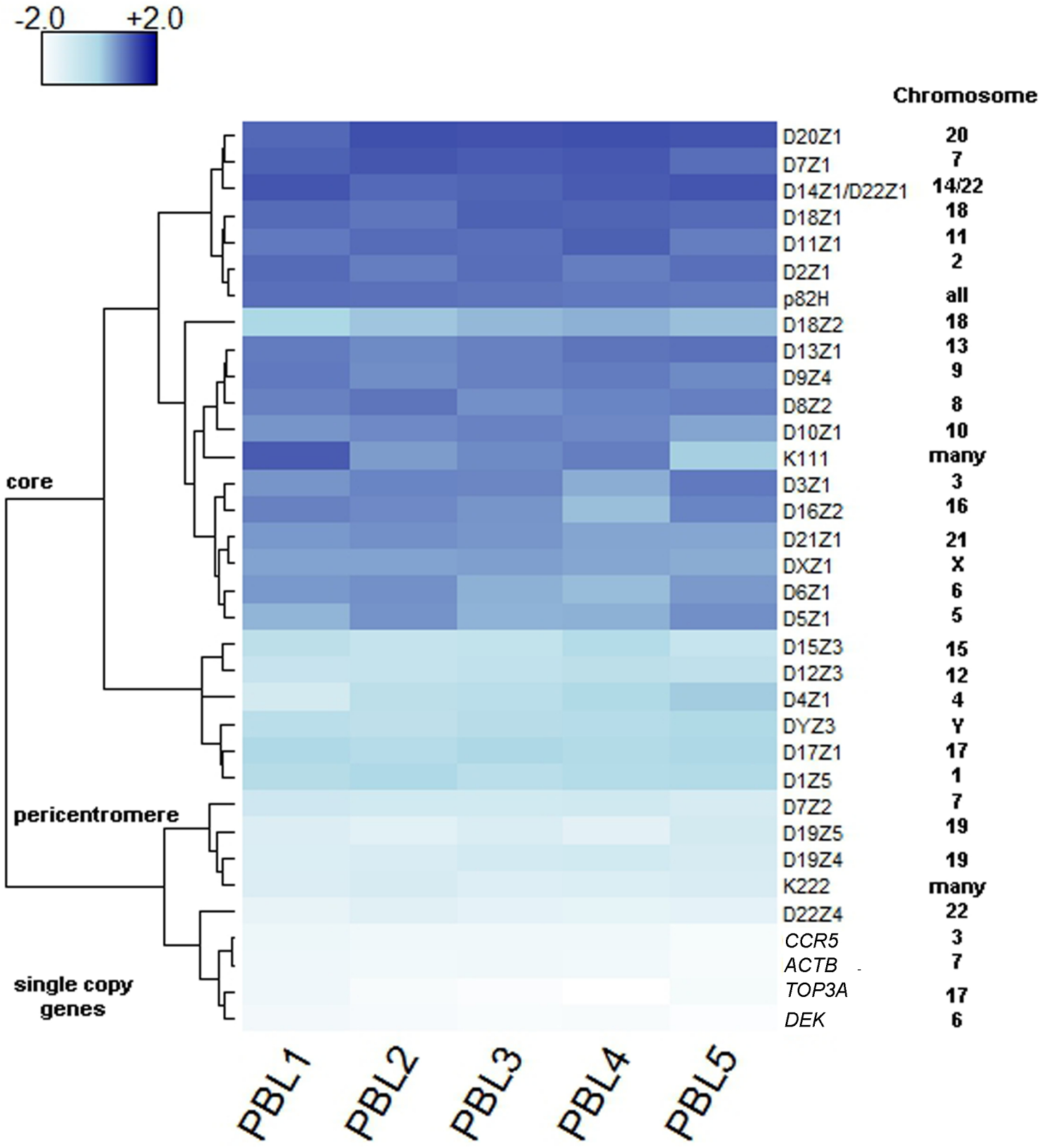
-2.0
+2.0
Chromosome
D20Z1 20
D7Z1 7
D14Z1/D22Z1 14/22
D18Z1 18
D11Z1 11
D2Z1 2
p82H all
D18Z2 18
D13Z1 13
D9Z4 9
D8Z2 8
D10Z1 10
K111 many
D3Z1 3
D16Z2 16
D21Z1 21
DXZ1 X
D6Z1 6
D5Z1 5
D15Z3 15
D12Z3 12
D4Z1 4
DYZ3 Y
D17Z1 17
D1Z5 1
D7Z2 7
D19Z5 19
D19Z4 19
K222 many
D22Z4 22
CCR5 3
ACTB 7
TOP3A 17
DEK 6
core
pericentromere
single copy genes
PBL1
PBL2
PBL3
PBL4
PBL5

Figure S6. Variation in the size of α-repeat arrays in individual humans. Heat map representing the abundance of α-repeats in each centromeric array and the number of pericentric retroviruses K111 and K222 (right, Y-axis) assessed by qPCR of 50 ng of DNA obtained from the human peripheral blood lymphocytes (PBLs) of five Caucassian individuals (X-axis). Note that D13Z1 and D21Z1 assays are not optimized using LNA primers in these experiments. The chromosomal location of each centromeric repeat is indicated on the right. No significant gene variation was seen in the single copy *TOP3A*, *CCR5*, *DEK*, or *ACTB* genes. The α-repeat p82H is present in the centromeres of all human chromosomes. The retroviruses K111 and K222 are present in the pericentromeres of fifteen and nine human chromosomes, respectively. The intensity of the heat map is depicted by the log base 2 Z-score of each α-repeat shown in the color gradient on the left. Hierarchical clustering trees were created to represent the content/size of every α-repeat array (tree). The tree splits into two main branches, one indicating arrays in centromere "cores" and a second with multiple-copy K111 and K222 sequences represented in higher abundance along with "pericentromere" arrays. Blue indicates higher copy numbers and white indicates lower copy numbers or single copy genes. Note that K111 clusters with the centromeric arrays, perhaps due to the fact it exists in so many copies, but also as it might indeed reside at the proximal pericentromere border as noted in the text. To specifically detect the core array D7Z2 in Chr 7 we elevated the primer annealing temperature, which led to substantially less detection of the monomers of this array. Therefore, the quantities of alphoid repeats estimated by PCR here clusters to arrays of smaller size, such as pericentromeric sequences.