



Additional File 4. The 'repeater' sequences and 'non-repeater' sequences represented by 60 dimension vector in the 3-dimensions scattered plot (3 dimensions \* 20 subplot, 3 dimensions are separately represent number, average number, variance). A, B, C, D: 3 Dimensions about A, G, C, T. E, F, G, H: 3 Dimensions about AA, AG, AC, AT. I, J, K, L: 3 Dimensions about GA, GG, GC, GT. M, N, O, P: 3 Dimensions about CA, CG, CC, CT. Q, R, S, T: 3 Dimensions about TA, TG, TC, TT. The blue and yellow dots represent known 'repeater' sequences and 'non-repeater' sequences respectively.