

## Duke/UNC/UT/EBI ENCODE insulator validation assays

The insulator enhancer-blocking assay was performed as previously described (Bell et al., 1999, Xi et al., 2006). Briefly, the beta-globin DNaseI HS2 site, which is a known enhancer element (Tuan et al., 1989), was cloned upstream of a NeoR gene. To test for insulator activity, DNA segments (**Table 1**) were amplified from human genome DNA by PCR, and cloned between the enhancer and NeoR gene. The previously described chicken insulator was used as a positive control (Chung et al., 1997). Test regions were divided into 4 categories in that they showed significant signals in:

- i) DNase-seq, FAIRE-seq and CTCF ChIP-seq assays and contained a CTCF motif
- ii) DNase-seq, FAIRE-seq, and CTCF ChIP-seq assays, but do not contain a known CTCF motif
- iii) DNase-seq and FAIRE-seq, but not CTCF ChIP-seq
- iv) Not detected by any assay

All plasmids were purified from three independent bacterial cultures, linearized, and each DNA prep was electroporated independently into K562 cells. Each electroporation was plated in triplicate (total of nine experiments per plasmid). The next day, cells were transferred to soft agar media containing G418. After 16 days, plates were scanned and colonies were counted. The enhancer blocking value =  $-\log_2 [(\text{\#colonies produced from the test site vector})/(\text{\#colonies produced from construct using chicken insulator positive control})]$ . Negative values indicate less enhancer-blocking activity than the standard, while positive values indicate enhancer-blocking activity more potent than the chicken insulator. Randomly selected regions were heavily enriched for negative values (**Table 1**).

### References:

Bell AC, West AG, Felsenfeld G (1999) The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98: 387-396.

Xi H, Shulha HP, Lin JM, Vales TR, Fu Y, et al. (2007) Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet* 3: e136.

Tuan DY, Solomon WB, London IM, Lee DP (1989) An erythroid-specific, developmental-stage-independent enhancer far upstream of the human “beta-like globin” genes. *Proc Natl Acad Sci U S A* 86: 2554–2558.

Chung JH, Bell AC, Felsenfeld G (1997) Characterization of the chicken beta-globin insulator. *Proc Natl Acad Sci U S A* 94: 575-580.

Table 1. Regions tested for insulator activity (hg19 coordinates)

Name	chr	start	stop	forward_primer	reverse_primer	insulator_value
<b>CTCF+DNaseI+FAIRE, no CTCF motif</b>						
Ins009	chr11	116602796	116604490	AAGGGGCAGGAACAAGAAGT	CTTGGGAAAAGGGGGATAA	4.73
Ins040	chr6	41559016	41560701	CACCAGGATGTTTCGCCTATT	CCAGCCCAGATGCTTCTTAC	0.43
Ins008	chr11	116561975	116563219	AGGGTCTCTACATTCATAACCT	CAATCCCTGCCTTCAGAGG	0.18
Ins002	chr11	2057880	2059108	GCTGCAGCTGGAGTTAGAGG	TACTGGAGGGGACGTTATG	-0.32
Ins014	chr14	99732310	99733733	CAAAAGCACATATATCCCCATC	GAGAGAAAACAGCCGTGAGC	-0.45
<b>CTCF+DNaseI+FAIRE+CTCF motif</b>						
Ins039	chr6	41336272	41337630	CTCTAACTCCCCATCCATGC	ATCCATGGGTTTCAGGTTTG	1.61
Ins023	chr2	220506552	220508135	AGAGATGTGCCTGGAACCAC	ACACAACAAACAGCCTCACG	1.11
Ins050	chr9	131893426	131895259	ACACCAGCTCTTGCTGACCT	GCCCTGACCTAGAGAACACCT	1.01
Ins001	chr1	151313091	151314485	CAGGCTAGGGAGGAACACAA	AGCTTGATACCCTGGGCTTT	0.65
Ins044	chr6	41743150	41744428	GTACTIONCCTGGCCTGTCT	TGGTGCTTTTGGAACTGTTG	0.28
Ins019	chr19	54344846	54346662	GCAAATTGAGGCTTGGGTTA	GGCAATTTTGCCTCTGTTT	0.18
Ins041	chr6	41570278	41571579	GGCATTCTTCCTATGCCAGA	CTGCCTCACTCTGTCCTTCA	-0.97
Ins048	chr7	27199987	27201408	TTTTGCCTCTGGAACCATTC	CTAACAGAGGCAGGGCTGAG	-1.48
Ins057	chrX	153218423	153219625	AACAGCAGAGAAGGGCATGT	GCTCCTTTCCCAACACAGAG	-0.10
Ins035	chr5	131836469	131837891	AGAGGCAGCTCAAGTGATCC	GGCCACTCTCTGTTTTGC	-0.51
<b>DNaseI+FAIRE positive, no CTCF ChIP-seq signal</b>						
Ins026	chr21	34754952	34756986	GCCCTTGACTTCAAAAAGGA	TGGTCTAGGTGCACCCTGTT	-1.10
Ins034	chr5	131831797	131833364	ACCGTGATGTGGAGTCCTTC	TGCCTCGAACTCACCTACT	-0.35
Ins056	chrX	122898473	122900188	AAAACCCAAGTAGACCTGGTGA	TTTGACCTCAATGGCAGAAA	-0.91
<b>Negative region, no chromatin signal</b>						
Random001	chr10	55839104	55840559	GCCATCTGAAGCTGTGATGA	CATGCTCTGGGAACTGGTTT	-1.93
Random002	chr18	25848817	25850340	CCGACAATTTGCTGGAAGAT	CACAGTTCGGGGGAATTAGA	-1.71
Random003	chr18	25926751	25928003	TCTGTGGATGGAAGGGACTC	AGAAAGGAGCCCTGATGGTT	-1.77
Random004	chr21	34848378	34849778	AAAATTTTCTGTTCCAGCCTCT	GACAGCGCAGAGGTATGTCA	-2.14
Random005	chr14	100394995	100396428	AGCTGCTGGTCTCTGTAGGG	CAAAGGGTCAGTGTCCACCT	-2.09
Random006	chr16	25991858	25993315	CACCATCTCCTGTGGTTCT	CGCATGCACATCTCCTAGA	-2.10
Random007	chr11	2136021	2137252	GTGCTGCAGGAAGGAGTTTC	CCTGTGTCTTCCAGGCAGTT	-1.81