**Table S1.** Total number of observed primate transposable elements (TEs) placed in 1-Mb windows, genome-wide for each lineage.

| TE | Human<br>(hg18) | Chimpanzee<br>(panTro2) | Orangutan<br>(ponAbe2) |
|---|---|---|---|
| Lineage-specific AluY | 13999 | 12935 | 22564 |
| HC[a] AluY | 18447 | - | - |
| HCO[b] AluY | 49904 | - | - |
| AluS | 303871 | - | - |
| AluJ | 137376 | - | - |
| | | | |
| Lineage-specific L1 | 10980 | 20300 | 47056 |
| HC L1PA | 18552 | - | - |
| HCO L1PA | 18458 | - | - |
| L1PB | 12701 | - | - |
| L1M | 307853 | - | - |

[a] Human-chimpanzee branch-specific elements

[b] Human-chimpanzee-orangutan branch-specific elements

Numbers of primate TEs are given after filtering events for segmental duplications, low sequence quality and coverage, and the pseudoautosomal boundaries on X and Y, and after filtering windows for outliers in regression models (see Methods for details).

**Table S2.** Genome features measured in 1-Mb windows.

| Feature | Significance | Calculation | Most relevant related mechanism shaping TE distribution |
|---|---|---|---|
| GC content | GC content reflects local base composition. | fraction of G and C bases per window (see Methods) | Integration preferences |
| L1 target site density | TTTTAA motif characteristic of target primed reverse transcription; from Cost et al (2002). | count per window | |
| Telomere-containing hexamer frequency[a] | TTAGGG motif characteristic of telomerase-dependent RNA retrotranscription and of non-canonical L1 integrations; from Nergadze et al. (2007) and Morrish et al. (2007). | ratio observed/expected frequency per window | |
| Replication timing* | Replication timing in S-phase of cell cycle; from Woodfine et al. (2004). | weighted average per window | |
| CpG island density[b] | CpG island density as a measure of local transcription activity, important to retrotransposition; from UCSC Genome Browser. | count per window | |
| CpG content | CpG dinucleotides are hypermutable due to methylation of cytosine; reflect local methylation status. | fraction of CpG bases per window | |
| Nucleosome-free region density* | Reflect local chromatin accessibility, measured by MNase cleavage; from Ozsolak et al. (2007). | count per window | |
| Germline-expressed region density* | Genes per window expressed (defined as average difference AD>200) in either testis germ cell or ovary tissue, to reflect transcriptionally active DNA; from Su et al. (2004). | count per window | |
| Lineage-specific divergence | Reflects simple sequence degradation; estimated in ARs with REV model implemented in HypHY (see Methods). | rate per bp per window | Simple sequence degradation |
| Small indel rates* | Reflects simple sequence degradation; from Kvikstad et al (2007). | rate (number of insertion or deletion events per aligned bp) per window | |
| Female recombination rate*[c] | Sex-specific recombination in the female germline; from Kong et al. (2002). | rate per window | Recombination |
| Male recombination rate* | Sex-specific recombination in the male germline;; from Kong et al. (2002). | rate per window | |
| Recombination hotspots | Computationally predicted hotspots of recombination using linkage- | count per window | |

| | | | |
|---|---|---|---|
| density* | disequlibrium among SNPs; from Myers et al. (2005). | | |
| Genome instability 13-mer frequency | CCNCCNTNNCCNC motif significantly associated with double-strand break formation and/or repair; from Myers et al (2008). | ratio observed/expected frequency per window | |
| Gene content[d] | Protein coding genes likely reflect local natural selection pressure. | fraction of bases per window | Natural selection |
| Most conserved elements density[e] | Regions likely to contain functional elements, and reflect local natural selection pressure; from Siepel et al. (2005). | count per window | |
| cis-NATs density* | cis-Natural Antisense Transcripts (cis-NATs) capable of gene expression regulation via RNA interference and/or transcriptional blocking; from Conley et al. 2008; (see Methods). | count per window | |
| Distance to telomere | Telomeres characteristically display heterogeneous rates and patterns of mutations (see text). | from midpoint coordinate of the window | |
| X chromosome indicator | Location of a window on X. | “1” for X; otherwise, “0” | Sex chromosome location |
| Y chromosome indicator[f] | Location of a window on Y. | “1” for Y ; otherwise, “0” | |

*Human-specific features unless noted otherwise.

[a] Frequency of a particular motif or K-mer was calculated as the ratio of observed/expected frequency, i.e. the number of K-mers per site of length K divided by the total of all possible K-mers, where K = length of motif, L= total length of sequence per window, $n_K$=number of K-mers, (L-K+1) is the number of sites of length K, and $1/(4^{\wedge K})$ is the number of all possible sequences of length K under the null assumption of equal base frequencies.

[b] CpG islands annotations from UCSC Genome Browser were obtained where available, except for orangutan (for orangutan this variable was not considered in the models).

[c] Human sex-specific recombination rates were used in chimpanzee models only, as it has been suggested that recombination rates are conserved across 1-Mb genomic scales at the evolutionary distance between human and chimpanzee (Myers et al 2005; Spencer et al 2006; Duret and Arndt 2008).

[d] Similar to Hellmann et al. (2005), gene content was calculated as fraction of bases per window located between CDS annotated start and stop coordinates of “Known genes”, “RefSeq genes”, “Ensembl genes” for human, chimpanzee, and orangutan, respectively.

[e] Most conserved elements were obtained from UCSC Genome Browser. For chimpanzee, human-orthologous regions were obtained from 28-way vertebrate alignments; for orangutan, most conserved elements were obtained from 8-way vertebrate alignments.

[f] The Y chromosome indicator variable is used in human and chimpanzee models only, the genomes for which Y chromosome sequence is available.

**Table S3**. Regression models for human Alu and L1 densities in 1-Mb windows genome-wide.

A. Alu elements

| | | Human-specific | HC[a] | HCO[b] | | |
|---|---|---|---|---|---|---|
| | Predictors | AluY | AluY | AluY | AluS | AluJ |
| Genome Landscape Model | Genome instability 13-mer frequency | 0.215 (+/31) | 0.103 (+/51) | 0.124 (+/61) | 0.148 (+/247) | 0.200 (+/221) |
| | L1 target site density | 0.122 (+/6) | 0.087 (+/39) | 0.267 (+/83) | 0.061 (+/89) | 0.076 (+/81) |
| | GC content | 0.073 (-/11) | 0.012 (-/5) | 0.025 (-/1) | 0.005 (-/2) | n.s.[c] |
| | Recombination hotspots density | 0.054 (-/3) | 0.008 (+/4) | 0.018 (+/7) | 0.007 (+/6) | 0.008 (+/6) |
| | Telomere-containing hexamer frequency | 0.017 (-/2) | 0.017 (-/8) | 0.006 (-/3) | n.s. | n.s. |
| | Distance to the telomere | 0.017 (+/2) | n.a.[d] | n.a. | 0.002 (-/4) | 0.002 (-/2) |
| | CpG island density | n.a. | n.s. | n.s. | 0.006 (-/7) | 0.006 (-/9) |
| | cis-NATs density | n.a. | 0.008 (+/4) | 0.006 (+/3) | 0.012 (+/15) | 0.007 (+/7) |
| | Germline-expressed region density | n.s. | 0.005 (+/2) | 0.012 (+/6) | 0.003 (+/7) | 0.003(+/2) |
| | Nucleosome-free region density | n.a. | 0.036 (-/17) | 0.010 (-/4) | 0.001 (-/2) | n.a. |
| | Most conserved elements density | n.s. | 0.011 (-/5) | n.a. | 0.012 (+/16) | 0.016 (+/19) |
| | Gene content | n.a. | 0.007 (-/3) | n.s. | n.s. | 0.004 (+/7) |
| | Female recombination rate | n.a. | n.a. | n.a. | 0.003 (+/5) | n.s. |
| | Male recombination rate | n.a. | n.s. | 0.006 (+/1) | n.a. | n.a. |
| | Replication timing | n.a. | n.s. | n.s. | n.s. | 0.003 (+/4) |
| | Human-specific divergence | n.a. | n.a. | n.a. | n.a. | n.a. |
| | Small indel rates | n.a. | n.a. | n.a. | n.a. | n.a. |
| | Interaction term[e] | (-/2) | n.s. | (-/13) | (-/4) | (+/1) |
| | Adjusted $R^2$ | 0.198 | 0.460 | 0.468 | 0.790 | 0.745 |
| X/Y/Autosome Model | X chromosome indicator | (-/5) | (-/6) | (-/10) | (-/9) | (-/5) |
| | Y chromosome indicator | (+/7) | (-/2) | (+/2) | (+/4) | (+/5) |
| | Adjusted $R^2$ | 0.015 | 0.009 | 0.015 | 0.018 | 0.013 |
| | Windows on X, Y, A[f] | 143, 15, 2567 | 145, 19, 2588 | 141, 8, 2548 | 140, 5, 2528 | 138, 6, 2544 |

[a] Human-chimpanzee branch-specific elements; [b] Human-chimpanzee-orangutan branch-specific elements; [c] n.s., not significant; [d] n.a., not applicable, i.e. not chosen by best subsets selection procedure; [e] GC content-L1 target site density interaction term: sign of beta coefficient and log10 of *p*-value are provided only. [f] Total windows for each chromosome type after filtering for outliers in regression models.

The RCVEs are indicated for each predictor significant after Bonferroni correction for multiple tests. RCVEs are ordered by rank of importance to Human-specific AluY model. Sign of beta coefficient and -log10 of Bonferroni-corrected *p*-value are provided in parentheses; in cases where both linear and quadratic terms are chosen by best subsets selection procedure, the results for beta coefficients and p-values for the most significant term, i.e. driving the relationship with the response, are presented here.

### B. L1 elements

| | | Human-specific | HC[a] | HCO[b] | | |
|---|---|---|---|---|---|---|
| | Predictors | L1 | L1PA | L1PA | L1PB | L1M |
| Genome Landscape Model | L1 target site density | 0.361 (+/132) | 0.070 (+/27) | 0.085 (+/30) | 0.058 (+/7) | 0.462 (+/102) |
| | GC content | 0.337 (-/3) | 0.063 (+/27) | 0.073 (+/22) | 0.057 (+/8) | 0.244 (+/24) |
| | Recombination hotspots density | 0.099 (-/31) | n.a. | n.a. | 0.055 (+/7) | n.s. |
| | Most conserved elements density | 0.057 (-/23) | 0.111 (-/44) | 0.064 (-/29) | 0.022 (-/5) | 0.050 (-/19) |
| | CpG island density | 0.012 (+/5) | 0.007 (+/3) | 0.009 (+/4) | 0.014 (+/3) | 0.015 (+/6) |
| | Telomere-containing hexamer frequency | 0.008 (+/3) | n.s. | n.a. | n.s. | 0.002 (+/2) |
| | Nucleosome-free region density | n.a. | 0.040 (+/5) | 0.008 (+/3) | 0.022 (+/3) | n.a. |
| | Genome instability 13-mer frequency | n.a. | n.a. | 0.024 (-/11) | 0.027 (-/4) | 0.228 (+/73) |
| | Distance to the telomere | n.s. | 0.007 (+/3) | n.s. | 0.012 (-/3) | 0.015 (-/6) |
| | Female recombination rate | n.a. | 0.010 (-/4) | 0.008 (-/3) | n.a. | n.a. |
| | Male recombination rate | n.s. | n.s. | 0.015 (+/7) | n.a. | n.s. |
| | Replication timing | n.s. | 0.020 (-/8) | n.s. | n.s. | n.s. |
| | Germline-expressed region density | n.a. | n.a. | n.s. | n.a. | 0.008 (-/3) |
| | cis-NATs density | n.s. | n.s. | n.a. | n.s. | n.s. |
| | Gene content | n.s.[c] | n.s. | n.a.[d] | n.a. | n.s. |
| | Human-specific divergence | n.a. | n.a. | n.a. | n.a. | n.a. |
| | Small indel rates | n.a. | n.a. | n.a. | n.a. | n.a. |
| | Interaction term[e] | (-/121) | (+/10) | n.s. | (-/1) | (-/37) |
| | Adjusted $R^2$ | 0.408 | 0.427 | 0.441 | 0.297 | 0.396 |
| X/Y/Autosome Model | X chromosome indicator | (+/14) | n.s. | (+/5) | (+/8) | (+/4) |
| | Y chromosome indicator | n.s. | n.s. | (+/2) | n.s. | n.s. |
| | Adjusted $R^2$ | 0.020 | n.s. | 0.007 | 0.010 | 0.004 |
| | Windows on X, Y, A[f] | 144, 19, 2610 | 143, 12, 2595 | 145, 13, 2582 | 146, 20, 2615 | 143, 10, 2522 |

[a] Human-chimpanzee branch-specific elements; [b] Human-chimpanzee-orangutan branch-specific elements; [c] n.s., not significant; [d] n.a., not applicable, i.e. not chosen by best subsets selection procedure; [e] GC content-L1 target site density interaction term: sign of beta coefficient and log10 of *p*-value are provided only. [f] Total windows for each chromosome type after filtering for outliers in regression models.

The RCVEs are indicated for each predictor significant after Bonferroni correction for multiple tests. RCVEs are ordered by rank of importance to Human-specific L1 model. Sign of beta coefficient and -log10 of Bonferroni-corrected *p*-value are provided in parentheses; in cases where both linear and quadratic terms are chosen by best subsets selection procedure, the results for the most significant term, i.e. driving the relationship with the response, are presented here (Human-specific L1 density model is an exception to this procedure, for which the negative quadratic term was chosen as added variable plots indicate a strong negative trend, despite the slightly larger p-value of this term).

**Table S4.** Primate lineage-specific regression models for Alu and L1 densities in 1-Mb windows, genome-wide for the subset of genome features in common to all primate species.

A. Alu elements

| | Predictors | Human-specific AluY | Chimpanzee-specific AluY | Orangutan-specific AluY |
|---|---|---|---|---|
| Genome Landscape Model | Genome instability 13-mer frequency | 0.208 (+/30) | 0.161 (+/55) | 0.132 (+/76) |
| | GC content | 0.116 (-/15) | 0.040 (-/8) | 0.076 (-/41) |
| | Distance to the telomere | 0.045 (+/5) | n.s.[a] | n.s. |
| | L1 target site density | 0.023 (+/1) | 0.063 (+/1) | 0.029 (-/10) |
| | Telomere-containing hexamer frequency | 0.019 (-/3) | 0.033 (-/12) | 0.006 (-/4) |
| | Most conserved elements density | n.s. | 0.006 (-/2) | 0.020 (-/10) |
| | CpG island density | n.s. | n.s. | n.a.[b] |
| | Gene content | n.a. | n.a. | n.a. |
| | Lineage-specific divergence | n.a. | n.a. | n.a. |
| | Female recombination rate | n.s. | n.s. | - |
| | Male recombination rate | n.a. | n.s. | - |
| | Interaction term[c] | n.s. | n.s. | n.s. |
| | Adjusted $R^2$ | 0.195 | 0.368 | 0.505 |
| X/Y/Autosome Model | X chromosome indicator | (-/2) | (+/3) | (-/11) |
| | Y chromosome indicator[d] | (+/9) | (+/5) | - |
| | Adjusted $R^2$ | 0.014 | 0.009 | 0.015 |
| | Windows on X, Y, A[e] | 143, 14, 2565 | 146, 7, 2611 | 142, 2620 |

[a] n.s. – non-significant

[b] n.a. – not applicable, i.e. not chosen by best subsets selection procedure

[c] GC content-L1 target site density interaction term: sign of beta coefficient and log10 of *p*-value are provided only.

[d] The Y chromosome indicator variable is used in human and chimpanzee models only, the genomes for which Y chromosome sequence is available.

[e] Total windows for each chromosome type after filtering for outliers in regression models; Orangutan genome windows on X and A, respectively.

The RCVEs are indicated for each predictor significant after Bonferroni correction for multiple tests. RCVEs are ordered by rank of importance to Human-specific AluY model. Sign of beta coefficient and -log10 of *p*-value are provided in parentheses; in cases where both linear and quadratic terms are chosen by best subsets selection procedure, the results for the most significant term, i.e. driving the relationship with the response, are presented here.

B. L1 elements

| | Predictors | Human-specific L1 | Chimpanzee-specific L1 | Orangutan-specific L1PA |
|---|---|---|---|---|
| Genome Landscape Model | L1 target site density | 0.362 (+/122) | 0.246 (+/124) | 0.077 (+/67) |
| | GC content | 0.353 (-/4) | 0.317 (-/8) | 0.080 (-/22) |
| | Most conserved elements density | 0.064 (-/22) | 0.054 (-/30) | 0.025 (-/26) |
| | Female recombination rate | 0.021 (-/5) | 0.012 (-/7) | - |
| | Male recombination rate | 0.014 (-/1) | 0.008 (+/4) | - |
| | CpG island density | 0.014 (+/5) | 0.006 (+/3) | n.a. |
| | Telomere-containing hexamer frequency | n.s. | 0.013 (+/8) | 0.006 (+/6) |
| | Distance to the telomere | n.s. | n.s. | 0.005 (+/4) |
| | Genome instability 13-mer frequency | n.a. | n.s. | 0.012 (-/8) |
| | Gene content | n.s. | n.a. | n.s. |
| | Lineage-specific divergence | n.a. | n.a. | n.a. |
| | Interaction term[c] | (-/120) | (-/134) | (-/74) |
| | Adjusted $R^2$ | 0.387 | 0.502 | 0.637 |
| X/Y/Autosome Model | X chromosome indicator | (+/16) | (+/7) | (+/16) |
| | Y chromosome indicator[d] | (+/2) | (+/9) | - |
| | Adjusted $R^2$ | 0.037 | 0.021 | 0.044 |
| | Windows on X, Y, A[e] | 143, 19, 2613 | 147, 21, 2610 | 149, 2628 |

[a] n.s. – non-significant
[b] n.a. – not applicable, i.e. not chosen by best subsets selection procedure
[c] GC content-L1 target site density interaction term: sign of beta coefficient and log10 of *p*-value are provided only.
[d] The Y chromosome indicator variable is used in human and chimpanzee models only, the genomes for which Y chromosome sequence is available.
[e] Total windows for each chromosome type after filtering for outliers in regression models; Orangutan genome windows on X and A, respectively.

The RCVEs are indicated for each predictor significant after Bonferroni correction for multiple tests. RCVEs are ordered by rank of importance to Human-specific L1 model. Sign of beta coefficient and -log10 of *p*-value are provided in parentheses; in cases where both linear and quadratic terms are chosen by best subsets selection procedure, the results for the most significant term, i.e. driving the relationship with the response, are presented here.

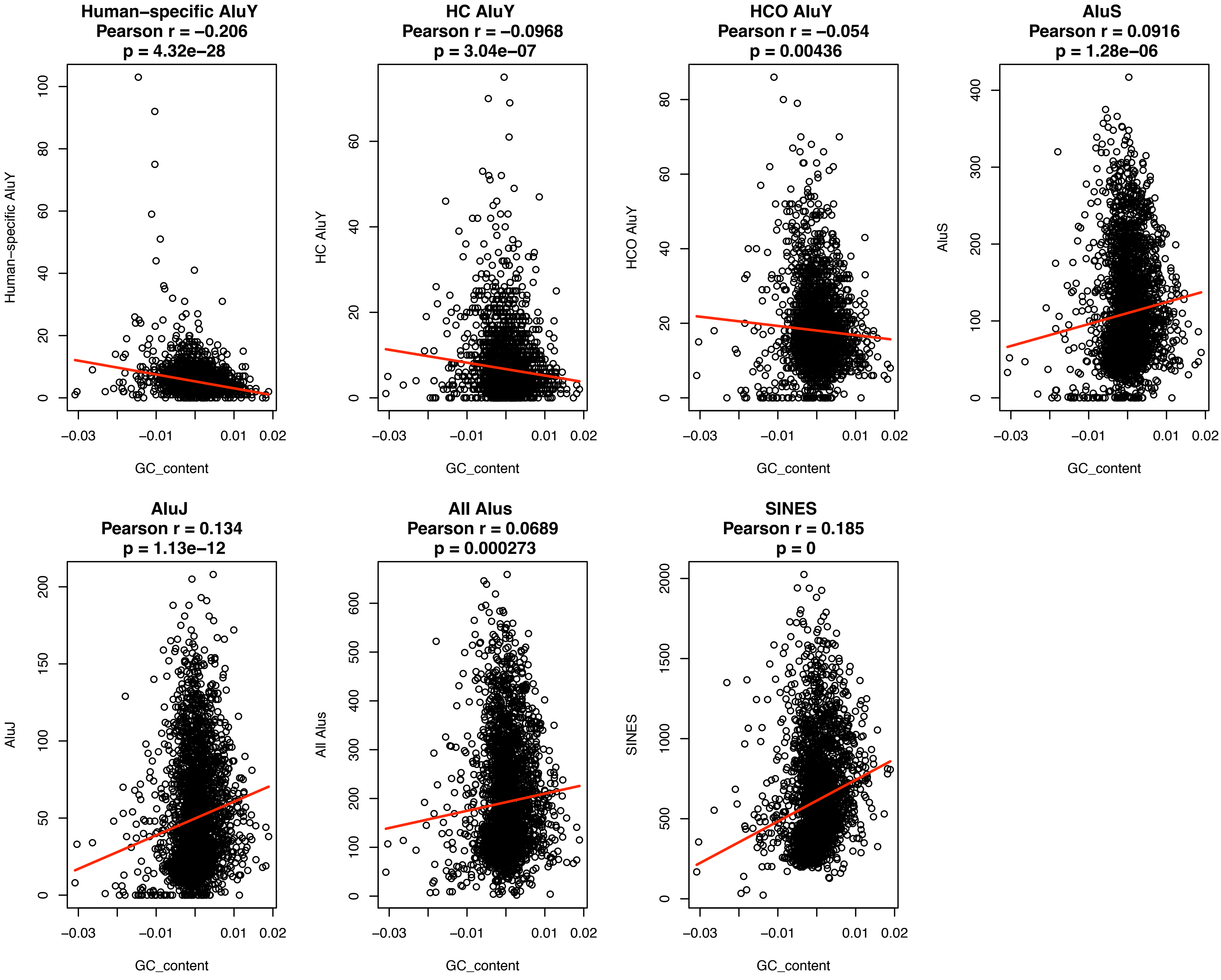
## SUPPLEMENTAL FIGURE LEGENDS

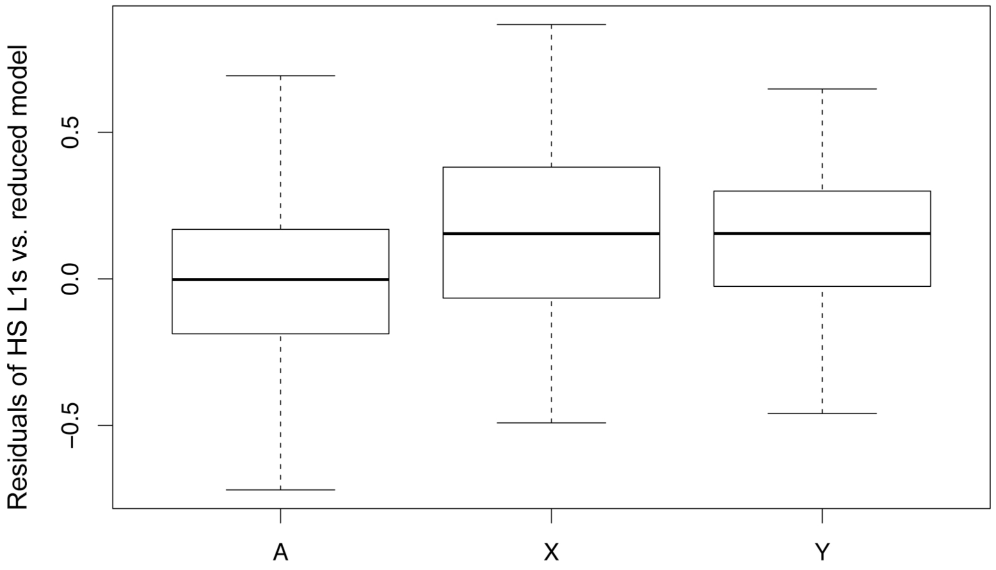
**Figure S1.** Pairwise correlations between human Alu densities and GC content observed in 1-Mb windows. Shown are scatterplots depicting pairwise relationships for each Alu subfamily investigated here: human-specific AluY, HC AluY, HCO AluY, AluS, AluJ. Additionally, pairwise correlations are plotted for all the above-mentioned Alus combined and for all annotated SINEs in the human genome. Pearson correlation coefficients and p-values are provided.

**Figure S2.** Human sex chromosome (X, Y) vs. autosome (A) distributions of corrected densities of L1s in 1-Mb windows genome-wide, excluding predictors of interest. Residuals from reduced multiple regression models (minus terms related to predictor of interest) represent densities corrected for local variation in the reduced set of genome landscape features significant to full models.

**Figure S3**. Orangutan sex chromosome vs. autosome distribution of observed and corrected densities for AluYs (A) and L1PAs (B) in 1-Mb windows genome-wide. (A) Observed densities are plotted for orangutan-specific AluY elements on the autosomes (yellow), and X (red). Residuals from genome-wide multiple regression models of AluYs represent densities corrected for local variation in genome landscape features. (B) Observed and corrected densities are plotted for orangutan-specific L1PAs.

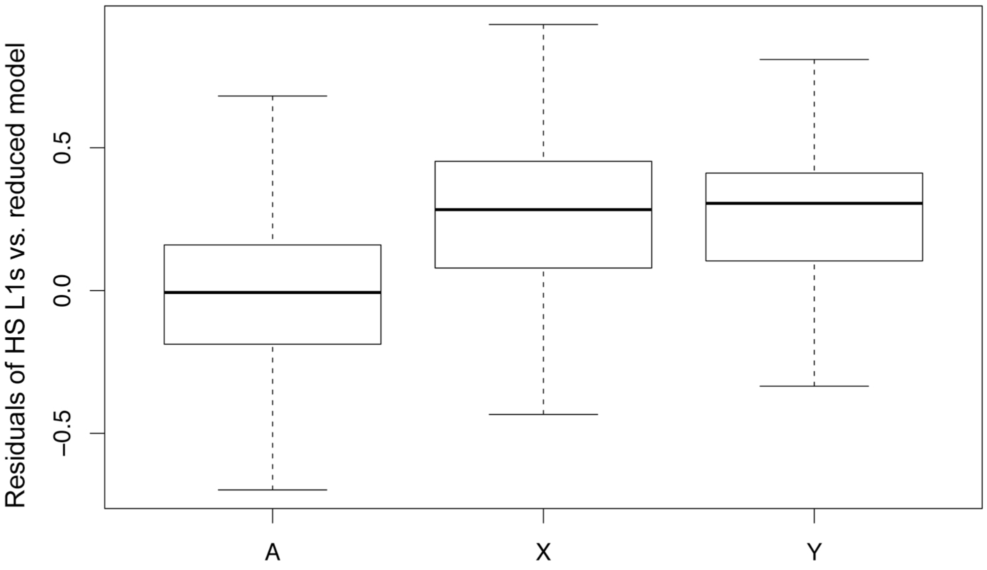
**Figure S4**. Chimpanzee sex chromosome vs. autosome distribution of observed and corrected densities of AluYs (A) and L1s (B) in 1-Mb windows genome-wide. (A) Observed densities are plotted for chimpanzee-specific AluY elements on the autosomes (yellow), X (red), and Y (blue) chromosomes. Residuals from genome-wide multiple regression models of AluYs represent densities corrected for local variation in genome landscape features. (B) Observed and corrected densities are plotted for chimpanzee-specific L1s.


Human−specific AluY
Pearson r = −0.206
p = 4.32e−28
HC AluY
Pearson r = −0.0968
p = 3.04e−07
HCO AluY
Pearson r = −0.054
p = 0.00436
AluS
Pearson r = 0.0916
p = 1.28e−06
AluJ
Pearson r = 0.134
p = 1.13e−12
All Alus
Pearson r = 0.0689
p = 0.000273
SINES
Pearson r = 0.185
p = 0
Human−specific AluY
HC AluY
HCO AluY
AluS
AluJ
All Alus
SINES
GC_content
−0.03
−0.01
0.01
0.02
0
20
40
60
80
100
150
200
300
400
500
600
1000
1500
2000



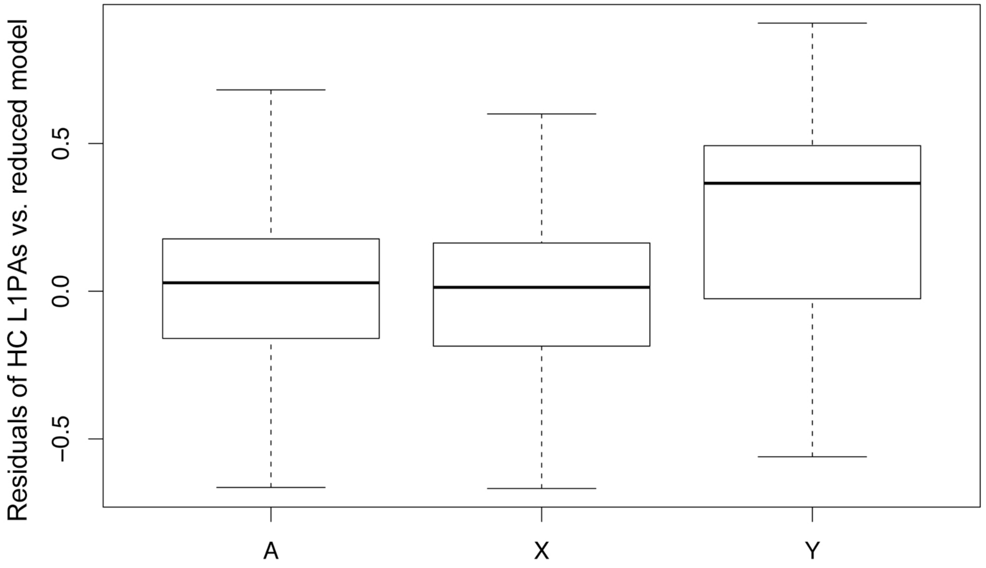
**Reduced model excluding most conserved elements density**

***p=2.95x10-12***



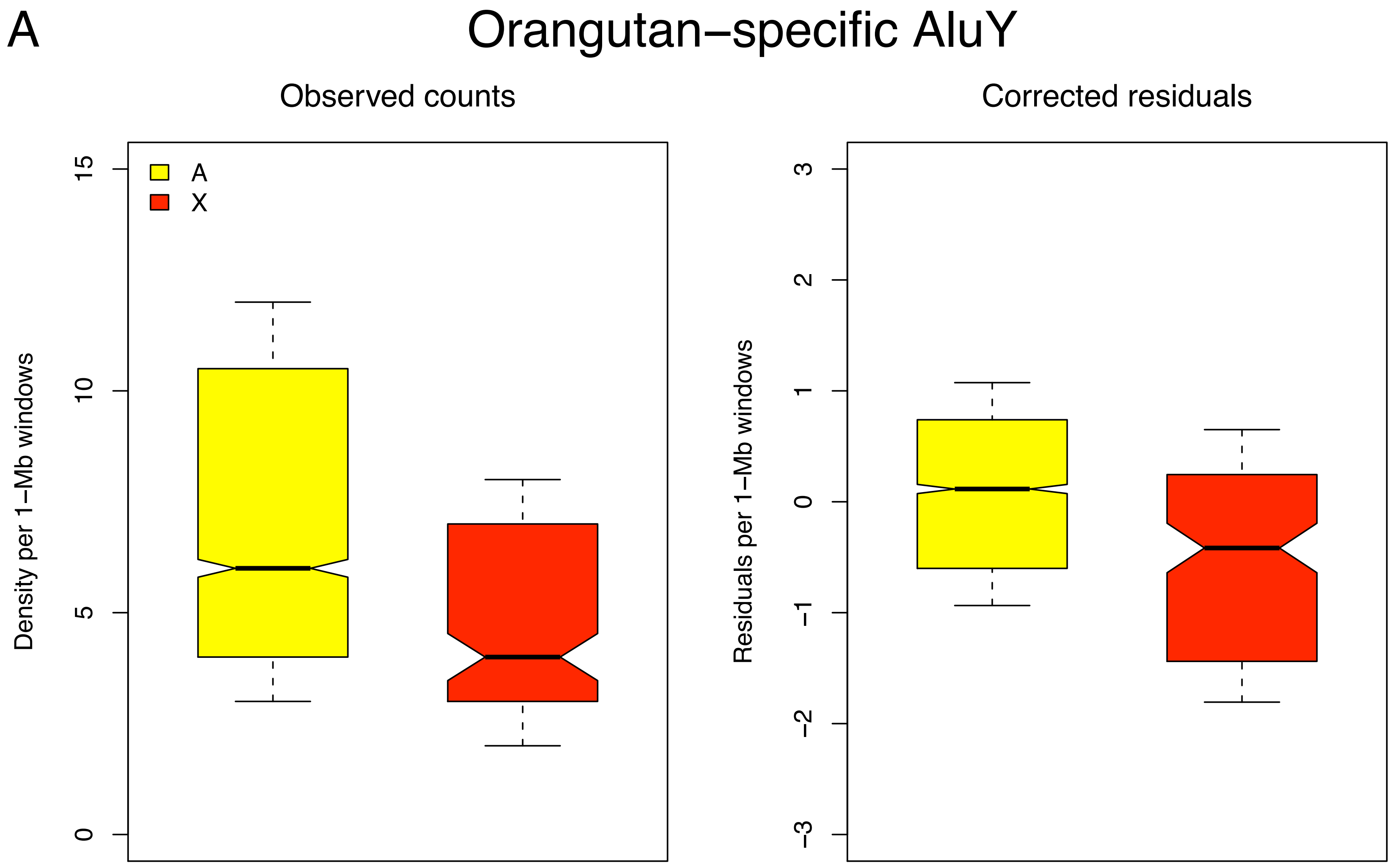
**Reduced model excluding recombination hotspot density**

***p<2x10-16***



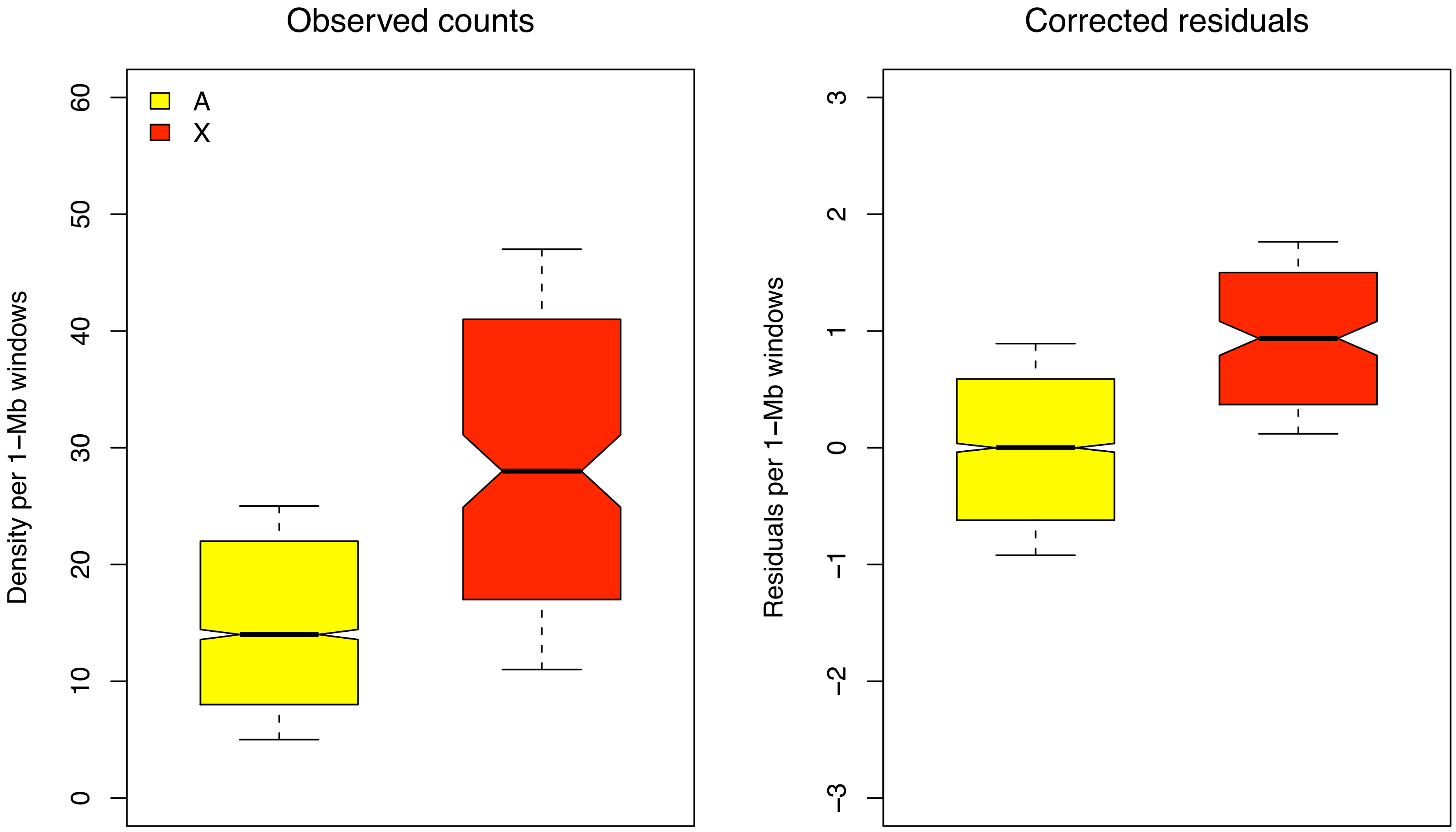
**Reduced model excluding most conserved elements density**

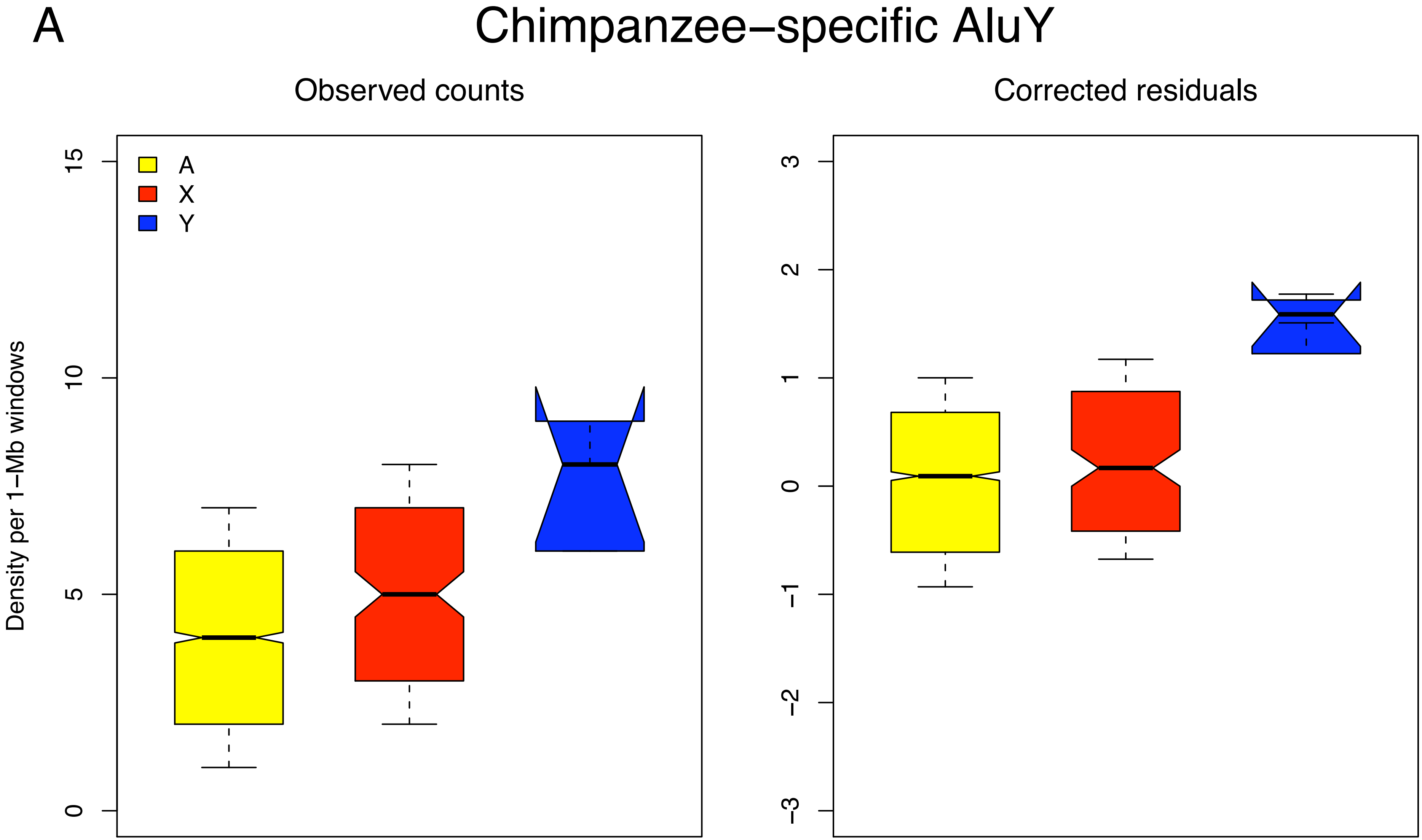
***p=0.0055***

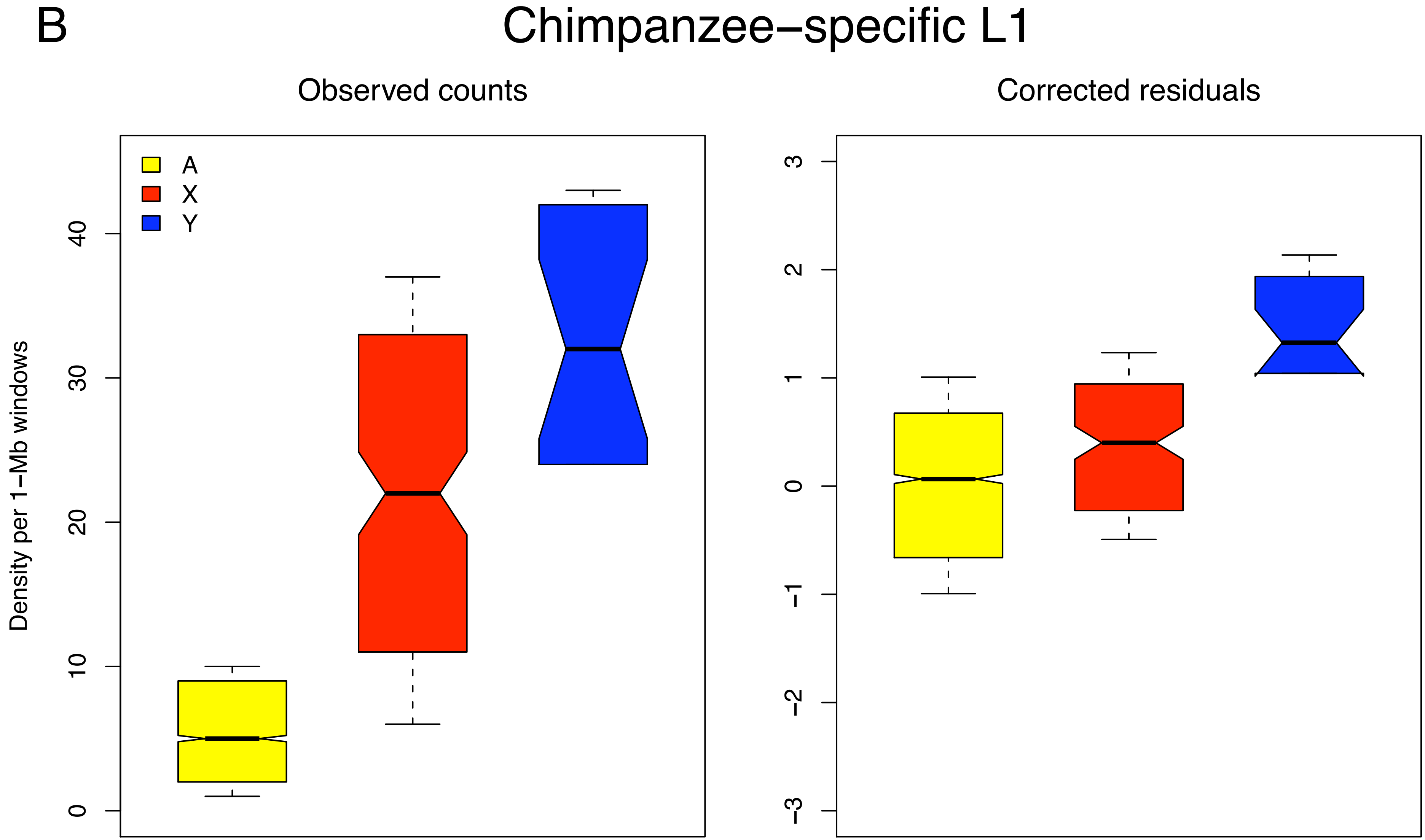

A
Orangutan−specific AluY
Observed counts
Corrected residuals
A
X
Density per 1−Mb windows
0
5
10
15
Residuals per 1−Mb windows
−3
−2
−1
0
1
2
3

B

Orangutan–specific L1PA




A
Chimpanzee–specific AluY
Observed counts
Corrected residuals
Density per 1–Mb windows
0
5
10
15
−3
−2
−1
0
1
2
3
A
X
Y


B
Chimpanzee−specific L1
Observed counts
Corrected residuals
A
X
Y
Density per 1−Mb windows
0
10
20
30
40
−3
−2
−1
0
1
2
3