Supporting Information

Yassour *et al.* 10.1073/pnas.0812841106

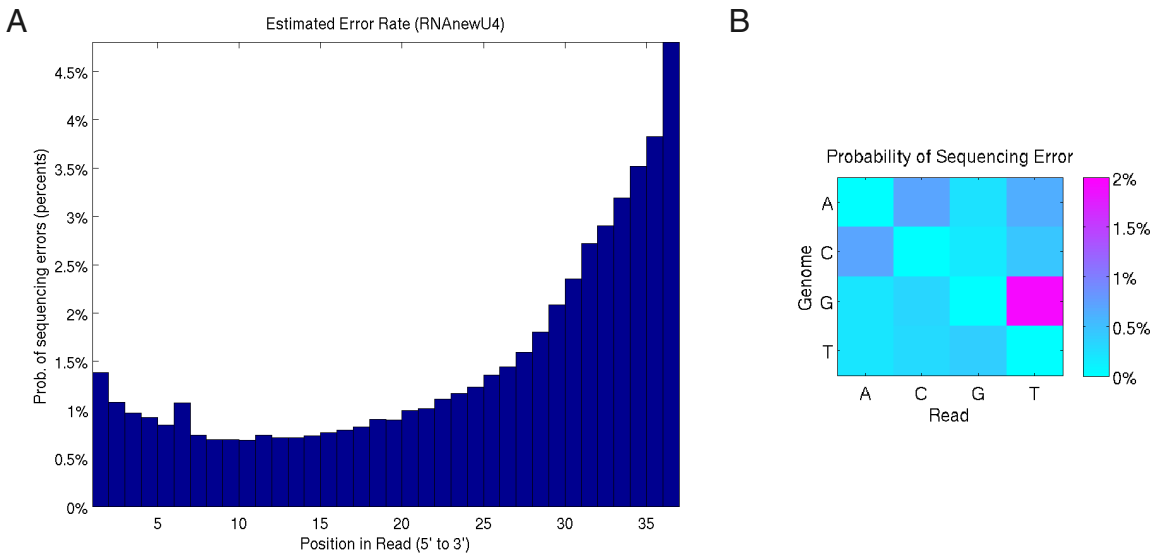


Fig. S1. Error model. (*A*) Estimated error rate for each position in the read. (*B*) The error rate of each specific error, averaged over all positions.

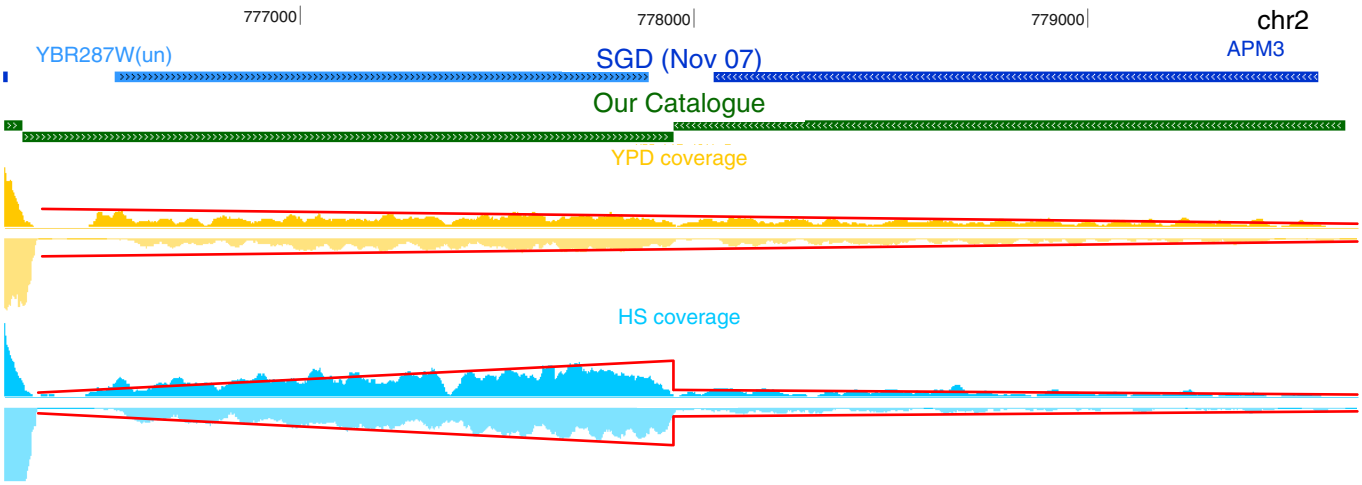


Fig. S2. Segmentation example. A visualization of the segmentation method applied on the locus chr2:776000–780000. In this example, the segmentation is almost impossible based on the YPD data alone, but when considering the HS data, it is very clear.

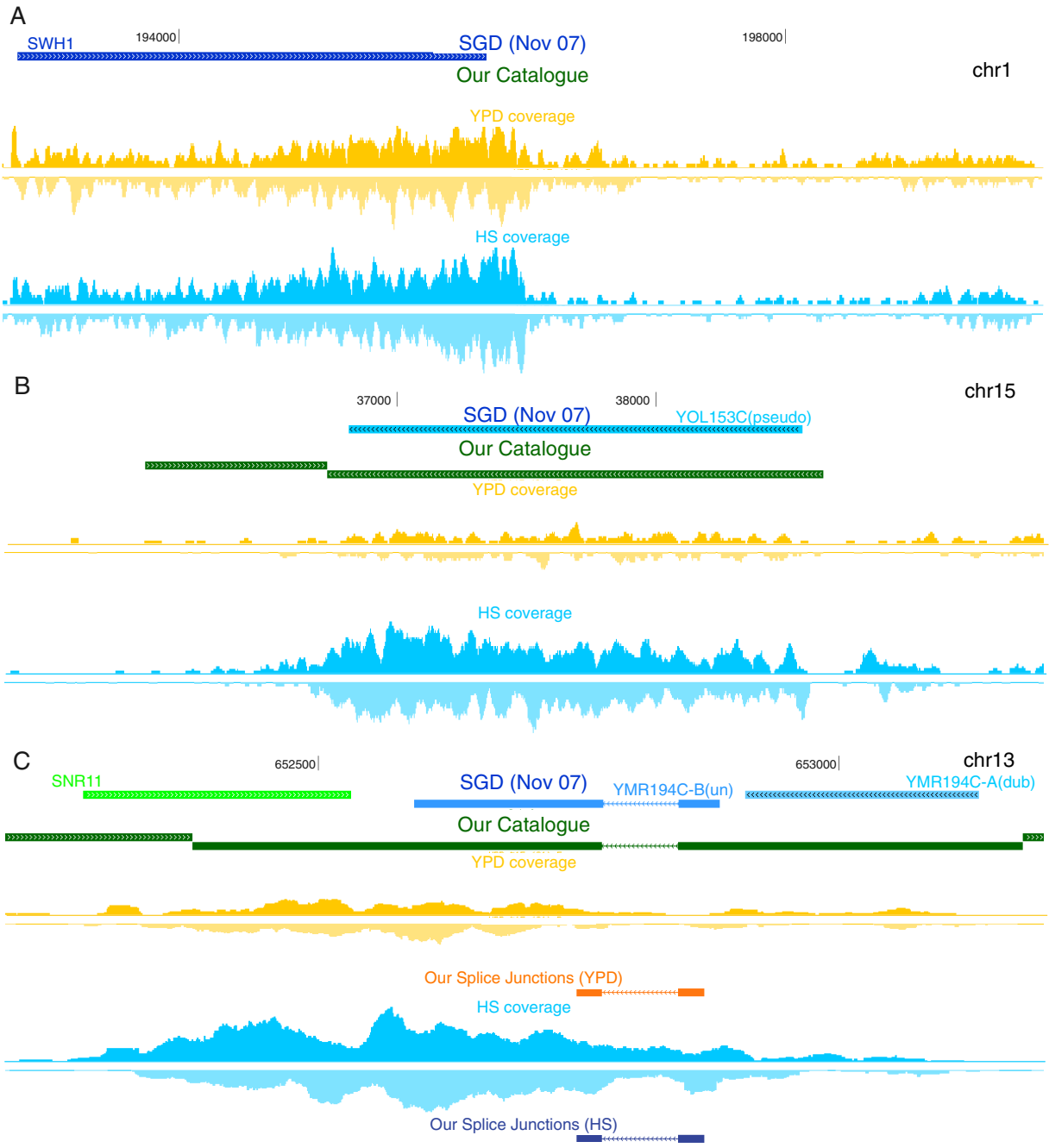
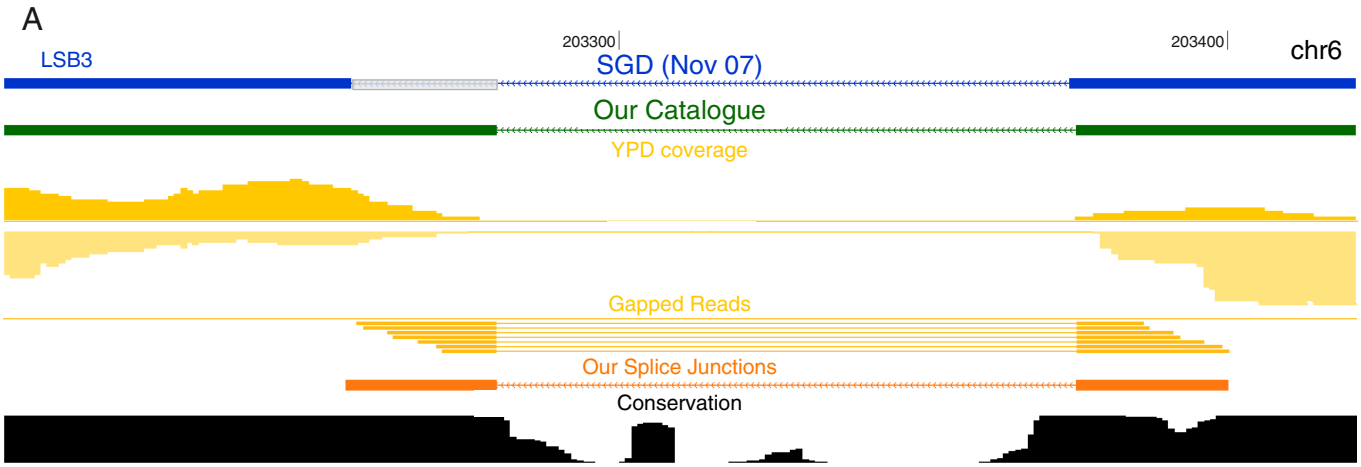


Fig. S3. Transcription validation. (*A*) A new transcribed element at chr1:196277–199970. (*B*) A transcribed pseudogene at chr15:36742–38650. (*C*) A novel transcription unit at the YMR194C locus that spans both a dubious ORF (YMR194C-B) and the gene YMR194C-A.



B

```
Spom|SPAPJ696.02          --MGLHNPLPSSLKSEC--KKAGKILTSFVDPRQTLGAQEVIPPSVLTNAKGL
Ylip|YALI0E05621g         --MGINNPVPRSLRSEC--RKAAKILASFVKPNQIFGQDMVIPPHVLQNAEGL
Calb|orf19.4127           --MGINNPIPRSLKSES--KKAAKILSSFIKPNQIAGPDQIIPPRILKNAKGL
Dhan|DEHA0C16093g         -------PVFR---------KAAKVLASFVKPNQFAGADQVIPPNVLKNAKGL
Scas|Scas607.11           -------MTSSYFLILT--RKAANVLASFVKPNQVLGADQIIPPDVLRRAKGL
Cgla|CAGL0A02145g        MIFCLTNPVPRGLANEA--QKAAKILEGFIDPRQAYGPDQVIPPSVLRNAKGL
Sbay|sbayc577-g5.1        --MGINNPIPRNLKSET--KKAAKILASFVKPNQVFGAEQVIPPDVLKRAKGL
Smik|smik268-g3.1         --MGINNPIPRNL-------KTAKILASFVKPNQVFGADQVIPPDVLKRAKGL
Scer|YHR016C              --MGINNPIPRSLKSET--KKAAKVLRSFVKPNQVFGADQVIPPYVLKRAKGL
Scas|Scas697.9            --MGINNPIPRSLSSETKTRKAAKILASFIKPNQVFGADQVIPPDVLKRAKGL
Cgla|CAGL0I08965g         --MGINNPIPRSLHSET--KKAAKILASFVKPNQVFGADQVIPPHVLKNAKGL
Sbay|sbayc606-g12.1       --MGFNNPIPRSLKSET--KKAANVLRSFVKPNQVFGADQVIPPYVLKKAKGL
Smik|smik318-g2.1         --MGINNPIPRSLKSET--KKAAKVLRSFVKPNQVFGADQVIPPYVLKKAKGL
Spar|spar434-g28.1        --MGINNPIPRNLKSET--KKAAKILASFVKPNQVFGADQVIPPDVLKRAKGL
Spar|spar37-g53.1         -------MIDSFDYFNH--RKAAKVLRSFVKPNQVFGADQVIPPYVLKRAKGL
Agos|AEL017W              --MGLNNPLPRSLTAET--KKAAKVLASFVKPLQMLGADEVIPPHVLKNAKGL
Klac|KLLA0A08360g         --MGINNPIPRSLKSET--KKAAKVLASFVKPNQVLSANDVIPPEVLKSAKGL

Scer|YFR024C-A (original)  --MGINNPIPRSLKSET----------  NFVKPNQVFGADQVIPPDVLKRAKGL

Scer|YFR024C-A (corrected) --MGINNPIPRSLKSET--  KKAAKILAS FVKPNQVFGADQVIPPDVLKRAKGL
```

Fig. S4. Splicing correction example. (*A*) In the gene LSB3, we find an intron that is shorter than reported by SGD [Cherry JM, *et al.* (1998) SGD: Saccharomyces Genome Database. *Nucleic Acids Res* 26:73–79]. The gray box represents the addition to the exon, according to our results. (*B*) The multiple sequence alignment of this region with the original and corrected annotation of the gene LSB3 [Wapinski I, Pfeffer A, Friedman N, Regev A (2007) Natural history and evolutionary principles of gene duplication in fungi. *Nature* 449:54–61]. It is clear that the added segment is highly conserved in other yeast species

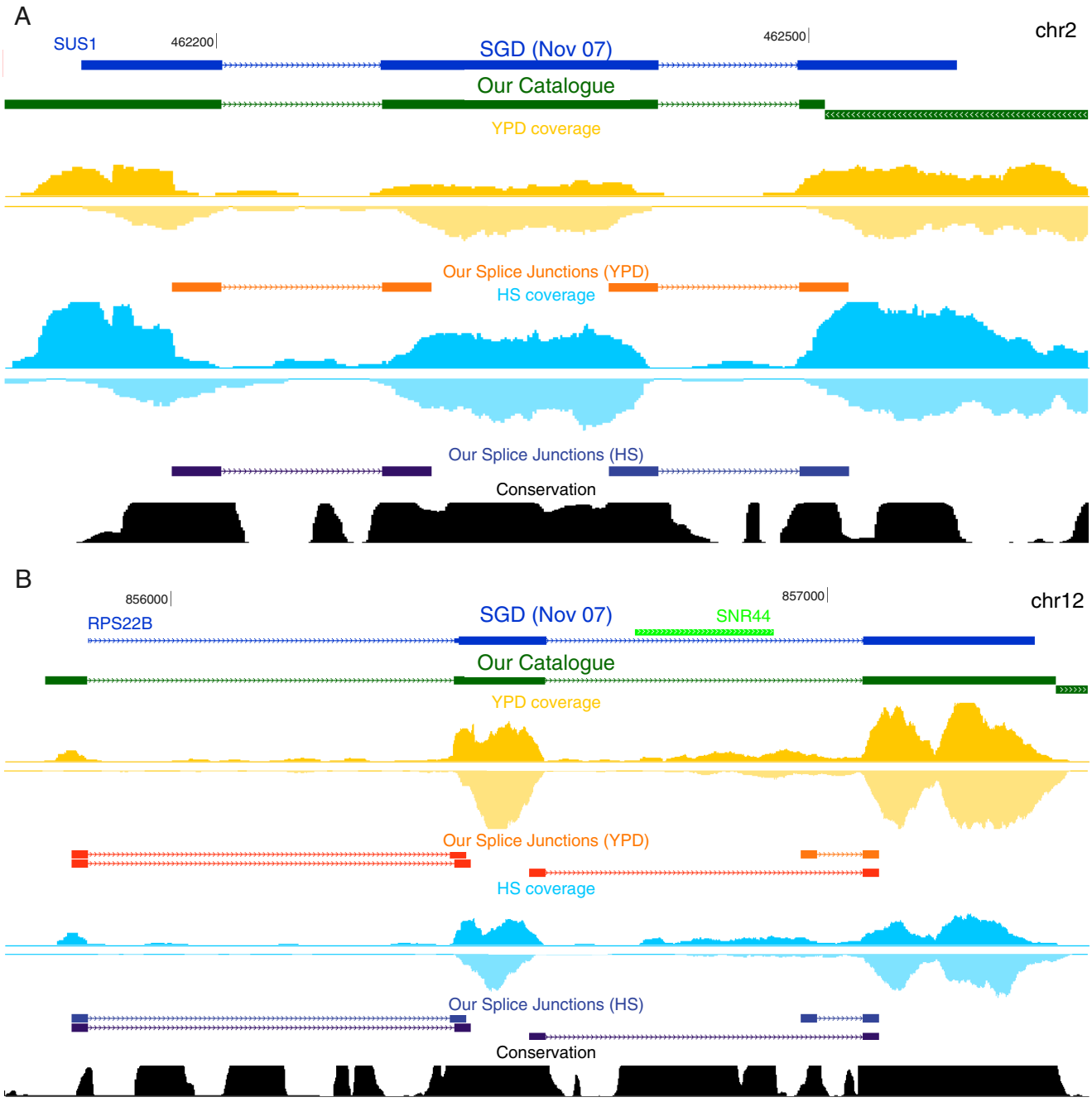


Fig. S5. Splicing validation. (*A*) Alternative splicing in the SUS1 gene, where, in addition to the 2 known introns, we also observe clear read-through at both junctions. Experimental validation confirms our predictions by revealing 3 bands, 2 bands consistent with just 1 intron spliced, and a stronger band consistent with both introns spliced out. (*B*) A previously uncharacterized intron from the end of the snoRNA, SNR44, to the acceptor site of its hosting intron, inside RPS22B.

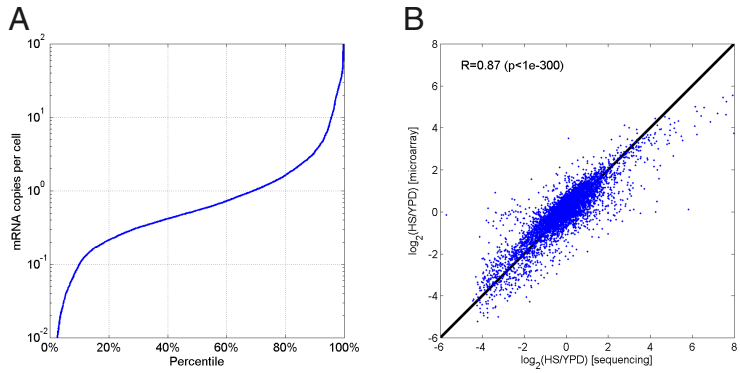


Fig. S6. Quantifying expression using sequencing. (*A*) Distribution of estimated mRNA copies per cell in YPD. Quantitative mRNA expression levels were estimated based on the density of reads along ORFs, with an estimate of 15,000 mRNA molecules per cell. (*B*) For each ORF, we computed the log2 ratio of HS and YPD (*x* axis), and compare it to its log2 ratio as measured by commercial 2-dye DNA microarrays (*y* axis).

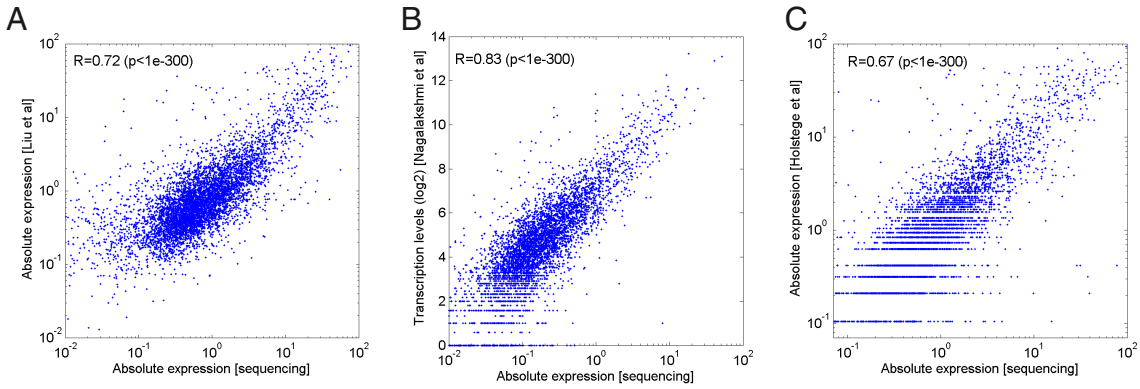


Fig. S7. Absolute expression comparison to previous studies [Nagalakshmi U, *et al.* (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320:1344–1349; Holstege FC, *et al.* (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95:717–728; Liu CL, *et al.* (2005) Single-nucleosome mapping of histone modifications in S. cerevisiae. *PLoS Biol* 3:e328.

Other Supporting Information Files

Dataset S1 (XLS)
Dataset S2 (XLS)
Dataset S3 (XLS)