Size selected NET-Seq reveals a conserved architecture of transcription units around yeast genes.

Jane Mellor[1], Shidong Xi[1], Struan Murray[1], and Philipp Lorenz[1]

[1]University of Oxford Department of Biochemistry

November 2, 2023

Abstract

Genomes from yeast to human are subject to pervasive transcription. A single round of pervasive transcription is sufficient to alter local chromatin conformation, nucleosome dynamics and gene expression, but experimentally is can be hard to distinguish from background signals. Size selected native elongating transcript sequencing (ssNET-Seq) was developed to precisely map transcription units (TU) independent of expression levels. RNAPII-associated nascent transcripts are fractionation into different size ranges before library construction. When anchored to the transcription start sites (TSS) of annotated genes, the combined pattern of the output metagenes define the expected reference pattern for a TU. Bioinformatic pattern matching to the reference identified 9542 TU in *Saccharomyces cerevisiae*, of which 47% are coding and 53% are non-coding. 3113 (33%) are newly identified unannotated non-coding TU. Anchoring all TU to the TSS or polyadenylation site (PAS) of annotated coding regions reveals distinctive architectures of linked pairs of divergent TU approximately 200nt apart. The Reb1 transcription factor is enriched 30nt downstream of the PAS only when an upstream (TSS-60nt) non-coding TU co-occurs with a downstream (TSS+150nt) coding TU and supports nucleosome depletion in the generation of the pervasive nascent transcriptome. The potential for extensive transcriptional interference is evident from low abundance unannotated TUs with variable TSS (median-240nt) initiating within a 500nt window upstream of, and transcribing over, the promoters of protein coding genes. This study confirms a highly interleaved yeast genome with different types of transcription units altering the chromatin landscape in distinctive ways, with the potential to exert extensive regulatory control.

Hosted file

`Xi et al Manuscript.docx` available at https://authorea.com/users/695004/articles/683893-size-selected-net-seq-reveals-a-conserved-architecture-of-transcription-units-around-yeast-genes

Figure 1

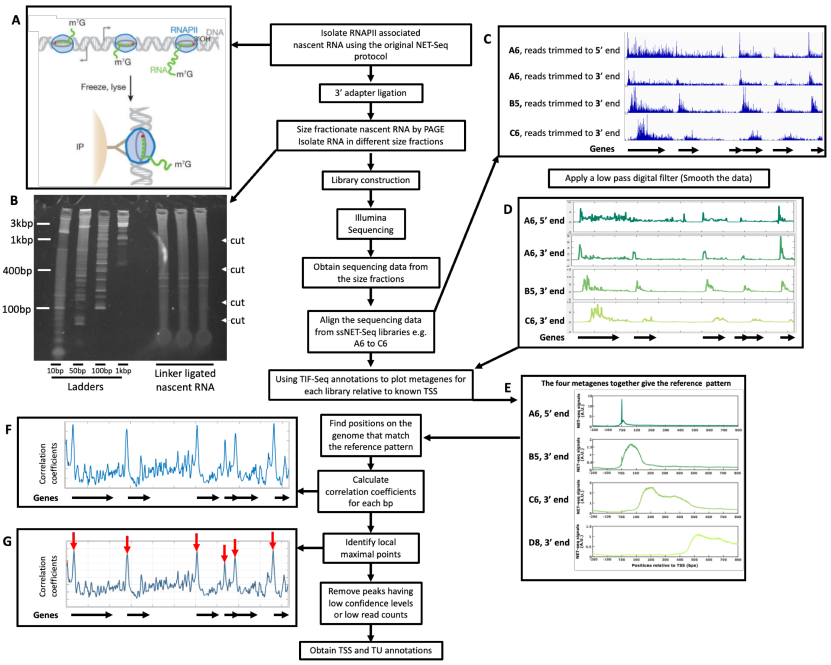


Figure 2

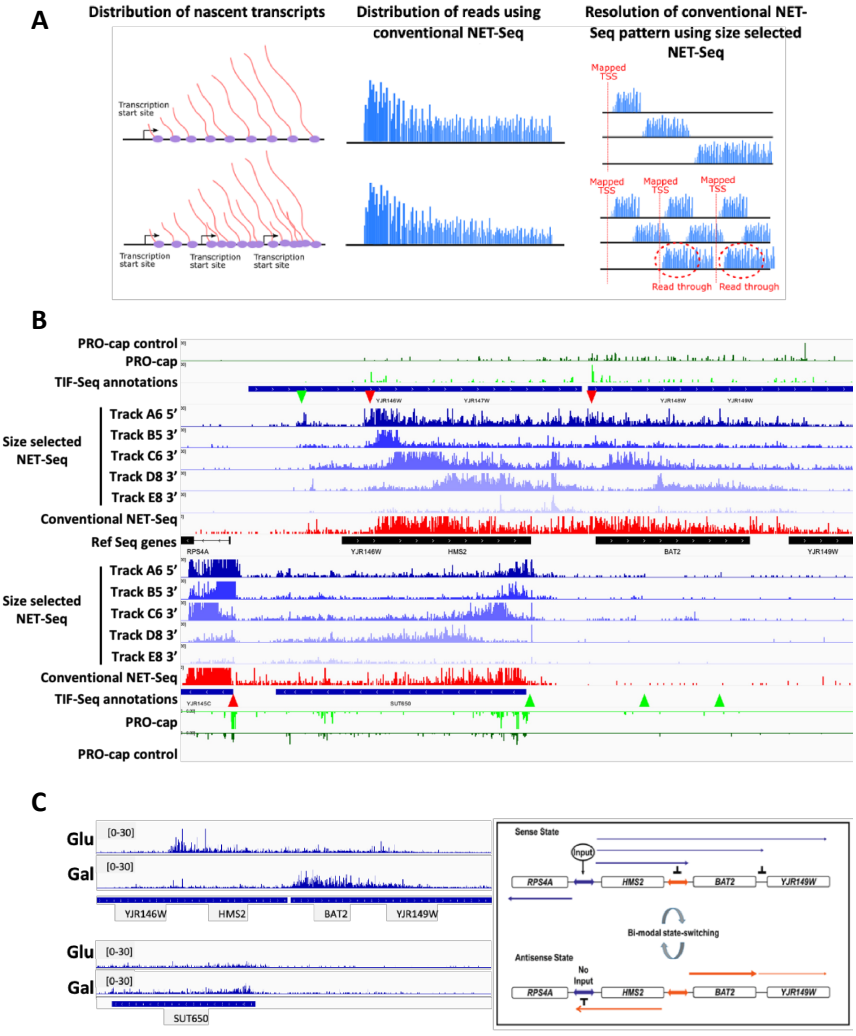


Figure 3

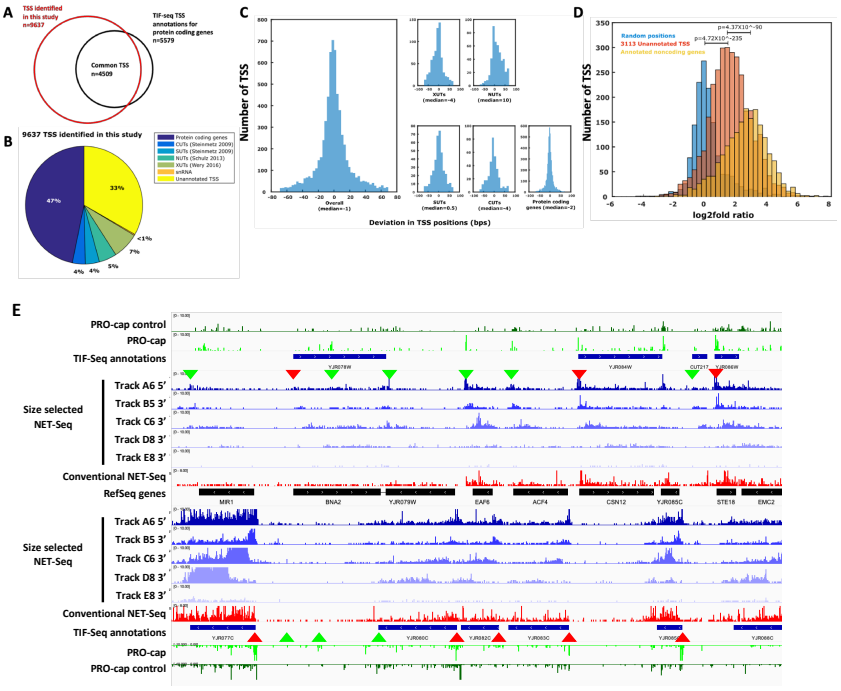


Figure 4

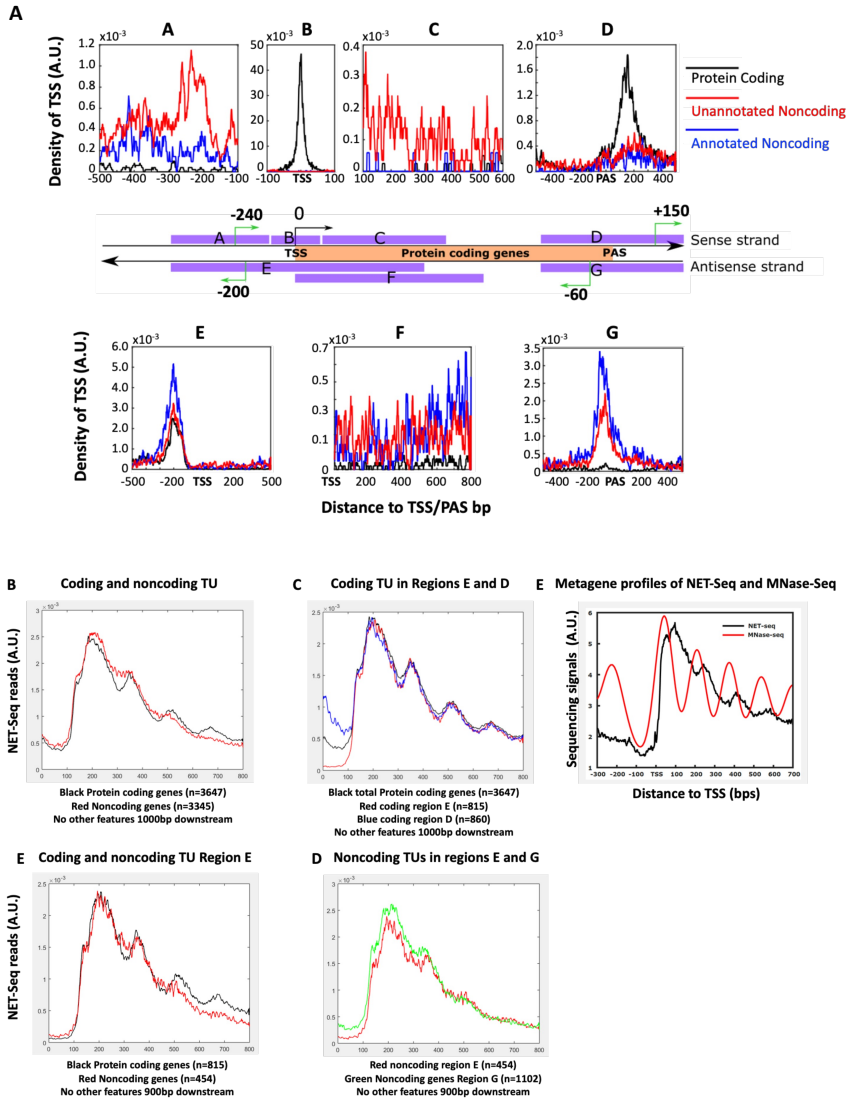
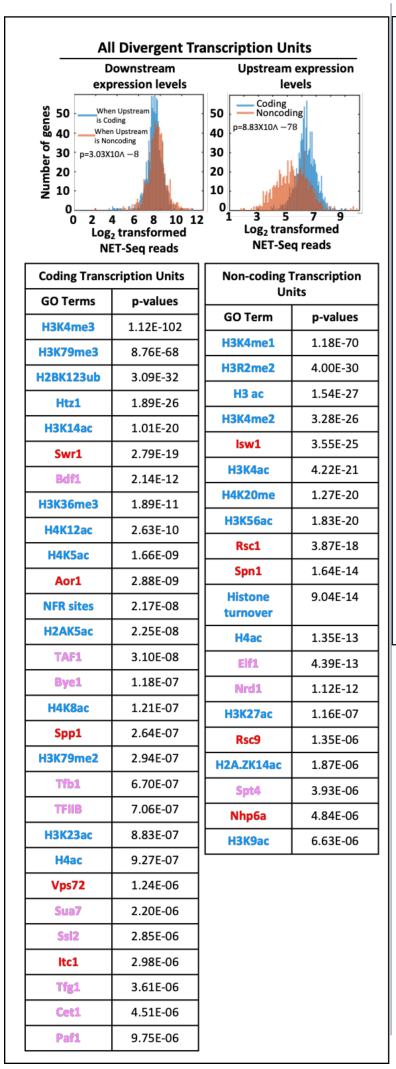
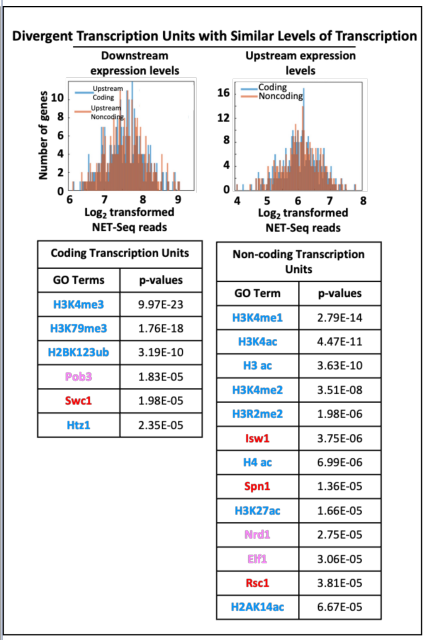


Figure 5



Coding Transcription Units	
GO Terms	p-values
H3K4me3	1.12E-102
H3K79me3	8.76E-68
H2BK123ub	3.09E-32
Htz1	1.89E-26
H3K14ac	1.01E-20
Swr1	2.79E-19
Bdf1	2.14E-12
H3K36me3	1.89E-11
H4K12ac	2.63E-10
H4K5ac	1.66E-09
Aor1	2.88E-09
NFR sites	2.17E-08
H2AK5ac	2.25E-08
TAF1	3.10E-08
Bye1	1.18E-07
H4K8ac	1.21E-07
Spp1	2.64E-07
H3K79me2	2.94E-07
Tfb1	6.70E-07
TFIIB	7.06E-07
H3K23ac	8.83E-07
H4ac	9.27E-07
Vps72	1.24E-06
Sua7	2.20E-06
Ssl2	2.85E-06
Itc1	2.98E-06
Tfg1	3.61E-06
Cet1	4.51E-06
Paf1	9.75E-06

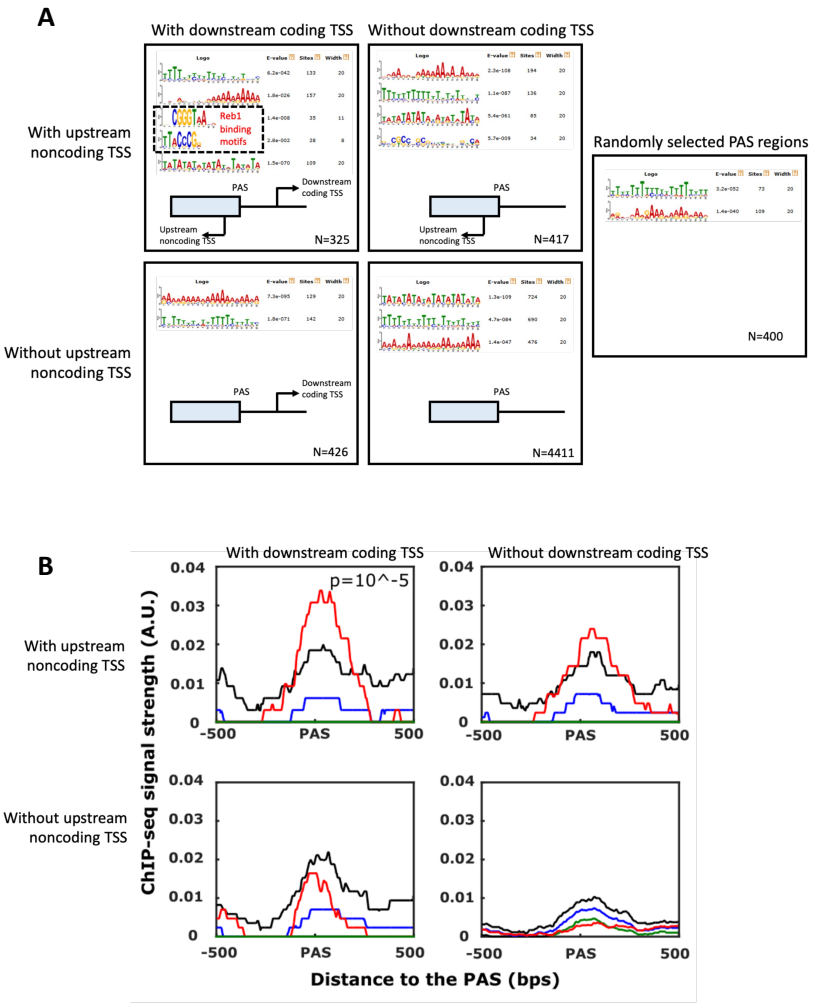
Non-coding Transcription Units	
GO Term	p-values
H3K4me1	1.18E-70
H3R2me2	4.00E-30
H3 ac	1.54E-27
H3K4me2	3.28E-26
Isw1	3.55E-25
H3K4ac	4.22E-21
H4K20me	1.27E-20
H3K56ac	1.83E-20
Rsc1	3.87E-18
Spn1	1.64E-14
Histone turnover	9.04E-14
H4ac	1.35E-13
Elf1	4.39E-13
Nrd1	1.12E-12
H3K27ac	1.16E-07
Rsc9	1.35E-06
H2A.ZK14ac	1.87E-06
Spt4	3.93E-06
Nhp6a	4.84E-06
H3K9ac	6.63E-06



Coding Transcription Units	
GO Terms	p-values
H3K4me3	9.97E-23
H3K79me3	1.76E-18
H2BK123ub	3.19E-10
Pob3	1.83E-05
Swc1	1.98E-05
Htz1	2.35E-05

Non-coding Transcription Units	
GO Term	p-values
H3K4me1	2.79E-14
H3K4ac	4.47E-11
H3 ac	3.63E-10
H3K4me2	3.51E-08
H3R2me2	1.98E-06
Isw1	3.75E-06
H4 ac	6.99E-06
Spn1	1.36E-05
H3K27ac	1.66E-05
Nrd1	2.75E-05
Elf1	3.06E-05
Rsc1	3.81E-05
H2AK14ac	6.67E-05

Figure 6



P-values	Downstream coding	Downstream noncoding	Downstream *annotated* noncoding	Downstream *unannotated* noncoding
Upstream coding	5.11 X 10^-05	1.32 X 10^-08	1.57 X 10^-06	0.000907585
Upstream noncoding	**1.51 X10^-311**	9.15 X 10^-12	-	-
Upstream *annotated* noncoding	**4.67 X 10^-159**	-	0.140176503	0.000390385
Upstream *unannotated* noncoding	**4.09 X 10^-96**	-	0.01240602	3.38 X 10^-06

Table 1 Correlations between an upstream (relative to the PAS) coding/noncoding TU and a downstream (relative to the PAS) coding/noncoding TU are tested by single sided Fisher's tests. The noncoding TUs are further divided into the annotated and unannotated noncoding TU. Combinations with a highly significantly positive correlation are highlighted in bold.