Transcription-dependent spreading of canonical yeast GATA factor across the body of highly expressed genes

Aria Ronsmans[1,3], Maxime Wéry[2,3], Camille Gautier[2], Marc Descrimes[2], Evelyne Dubois[1], Antonin Morillon[2,4] and Isabelle Georis[1,4]

[1] Metabolism of Model Microorganisms, Institut de Recherches Microbiologiques J.M. Wiame - Labiris, Laboratoire de Microbiologie, Université Libre de Bruxelles, avenue Emile Gryson 1, 1070 Brussels, Belgium

[2] ncRNA, epigenetic and genome fluidity, Institut Curie, PSL Research University, CNRS UMR 3244, Université Pierre et Marie Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France

[3] Equal contribution

[4] Corresponding authors: igeoris@ulb.ac.be, antonin.morillon@curie.fr

Summary

GATA transcription factors are highly conserved among eukaryotes and play roles in transcription of genes implicated in cancer progression and hematopoiesis. However, although their consensus binding sites have been well defined *in vitro*, the *in vivo* selectivity for recognition by GATA factors remains poorly characterized. Using ChIP-Seq, we identified the Dal80 GATA factor targets in yeast. Our data reveal Dal80 binding to a large set of promoters, sometimes independently of GATA sites. Strikingly, Dal80 was also detected across the body of promoter-bound genes, correlating with high, Dal80-sensitive expression. Mechanistic single-gene experiments showed that Dal80 spreading across gene bodies is independent of intragenic GATA sites but requires transcription elongation. Consistently, Dal80 co-purified with the post-initiation form of RNA Polymerase II. Our work suggests that GATA factors could play dual, synergistic roles during transcription initiation and elongation, promoting efficient remodeling of the gene expression program in response to environmental changes.

Introduction

In eukaryotes, gene transcription by RNA polymerase II (PolII) is initiated by the binding of specific transcription factors to double stranded DNA. The yeast transcription factors target regulatory regions called UAS or URS (for *U*pstream *A*ctivating/*R*epressing *S*equences), generally directly adjacent to the core promoter. The generated regulatory signals converge at the core promoter where they permit the regulation of PolII recruitment via the 'TATA box-binding protein' and associated general transcription factors (Hahn and Young, 2011; Levine and Tjian, 2003). The transcription factor binding sites are usually short sequences ranging from 8 to 20 bp (Zambelli et al., 2013). They are most often similar but generally not identical, differing by some nucleotides from one another (Zambelli et al., 2013), making it sometimes difficult to predict whether a given UAS will function as such *in vivo*.

GATA factors constitute a family of transcription factors highly conserved among eukaryotes and characterized by the presence of a DNA binding domain which consists of four cysteines (fitting the consensus sequence $CX_2CX_{17\text{-}18}CX_2C$) coordinating a zinc ion followed by a basic carboxy-terminal tail (Omichinski et al., 1993). *In vitro*, this conserved binding domain allows the recognition of consensus GATA DNA sequences (GATAA, GATAAG or GATTAG), found in the promoter of numerous genes, which was demonstrated in various organisms either with direct or indirect methods (Cunningham and Cooper, 1993; Horak et al., 2002; Omichinski et al., 1993; Stanbrough and Magasanik, 1996; Starich et al., 1998). Structure elucidation of GATA factor binding domains in complex with their DNA target suggest an important role for the basic tail in sequence specificity determination (Omichinski et al., 1993; Starich et al., 1998).

Since its discovery 40 years ago in chicken cells, the family of GATA factors was extended in human cells and represents master regulators of hematopoiesis and cancer (Bresnick et al., 2012). However, although approximately 7 million GATA motifs can be found in the human genome, the GATA factors occupy only 0.1–1% of them. Conversely, other regions are occupied by GATA factors despite lacking the consensus motif (Cheng et al., 2009; Tripic et al., 2009). In addition, GATA factors

can swap among them for the same motif and switch from active or repressive transcriptional activity. All these observations raised the main paradigm on understanding the selective recruitment and activity of GATA factors and what the determinants of their residency on the chromatin are (Bresnick et al., 2010; Grass et al., 2003).

In yeast, the family of GATA transcription factors contains over 10 members (Lowry and Atchley, 2000). Four of them are implicated in the regulation of *N*itrogen *C*atabolite *R*epression-(NCR-) sensitive genes, the expression of which is repressed in the presence of a preferred nitrogen source (glutamine, asparagine, ammonia) and derepressed when only poor nitrogen sources (e.g. proline, leucine, urea) are available (Magasanik and Kaiser, 2002). These key GATA factors involved in NCR signaling are two activators (Gln3 and Gat1/Nil1) and two repressors (Gzf3/Nil2/Deh1 and Dal80/Uga43) (Coffman et al., 1996; Coornaert et al., 1992; Cunningham and Cooper, 1991; Minehart and Magasanik, 1991; Rowen et al., 1997; Soussi-Boudekou et al., 1997). Moreover, the expression of *DAL80* and *GAT1* is also NCR-sensitive, which implies cross- and autogenous regulations of the GATA factors in the NCR mechanisms (Coffman et al., 1997; Georis et al., 2009, 2015; Soussi-Boudekou et al., 1997). Under nitrogen limitation, expression of *DAL80* is highly induced (Cunningham and Cooper, 1991), and Dal80 enters the nucleus where it competes with the two GATA activators for the same binding sites (André et al., 1995; Coffman et al., 1997; Cunningham et al., 1994). Although initially described as being active under nitrogen abundance (Rowen et al., 1997; Soussi-Boudekou et al., 1997), the Gzf3 repressor also localizes to NCR-sensitive promoters in conditions of activation (Georis et al., 2009).

In vitro, the Gln3 and Gat1 activators bind, presumably as monomers, to single GATA sequences (Cunningham et al., 1996), like their orthologous vertebrate counterparts. On the other hand, the Dal80 repressor can form homo-and heterodimers with Gzf3, thanks to the presence of an N-terminal leucine zipper (Svetlov and Cooper, 1998) and, consistently, Dal80 binds *in vitro* to two GATA sequences (Coffman et al., 1997; Cunningham and Cooper, 1993; Cunningham et al., 1994; Svetlov and Cooper, 1998). *In vivo*, GATA factor binding site recognition also appears to require

repeats of GATA motifs within promoters. Indeed, synthetic constructions containing a portion of the NCR-sensitive *DAL5* promoter have shown in a β-galactosidase assay that at least two copies of the consensus are required to support efficient gene activation. This led to the actual fuzzy definition of UAS_{NTR}, consisting in two GATA sites located sufficiently close to one another to present a binding platform for GATA factors (Bysani et al., 1991; Cooper et al., 1989; Rai et al., 1989).

In vivo, GATA ChIP analyses have been carried out genome-wide in yeast (Harbison et al., 2004; Lee, 2002; MacIsaac et al., 2006; Venters and Pugh, 2009). They allowed gaining insights into the GATA factor gene network through the identification of direct targets and characterization of the DNA elements required to render a gene NCR-sensitive. However, these studies were not performed in activating conditions, when all GATA factors are expressed, localized in the nucleus and active, so that the current list of GATA factor targets are likely to be underestimated. On another hand, bioinformatic analyses have shown that, since GATA sequences are short, they can be found almost everywhere throughout the genome. Therefore, based on the sole criteria of the presence of repeated GATA sequences in yeast promoters, a third of the yeast genes could hypothetically be NCR regulator targets (van Helden et al., 1998). Such GATA motif repetitions have been found in the promoter of 91 genes, inducible by GATA activators in absence of a good nitrogen source, supposed to be directly targeted by the GATA activators (Scherens et al., 2006). Nevertheless, the functionality of these supposed UAS still needs to be directly demonstrated *in vivo* (Hahn and Young, 2011).

Here, we provide the first genome-wide identification of Dal80 targets in yeast, in physiological conditions where Dal80 is fully expressed and active. Using a ChIP-Seq approach combined to a bioinformatic peak-calling procedure, we defined the exhaustive set of Dal80-bound promoters, which turned out to be much larger than anticipated. Our data indicate that at some promoters, Dal80 recruitment occurs independently of GATA sites. Strikingly, Dal80 was also detected across the body of a subset of genes bound at the promoter, globally correlating with high and Dal80-sensitive expression. Mechanistic single-gene experiments confirmed the Dal80 binding profiles, further indicating that Dal80 spreading across gene bodies occurs independently of intragenic GATA sites but

requires transcription elongation. Finally, co-immunoprecipitation experiments revealed that Dal80 physically interacts with both intitiating and elongating form of PolII. Our work suggests that in addition to their well-established role at the (pre)inititation stage, GATA factors could be involved in transcription elongation, promoting efficient remodeling of the gene expression program in response to environmental changes.

Results

Genome-wide identification of Dal80-bound promoters

In order to determine the genome-wide occupancy of a GATA factor in yeast, our rationale was to choose Dal80 as it is known to be highly expressed in derepressing conditions and forms chromosome foci when tagged by GFP (Distler et al., 2001). We thus grew yeast cells in proline-containing media and performed a ChIP-Seq analysis using a Dal80-Myc[13] tagged strain and the isogenic untagged strain, as a control (Fig. 1A). Dal80-bound regions were identified using a peak-calling algorithm (see Material & Methods). A promoter was defined as bound by Dal80 when >75% of the -350 to -100 region (relative to the downstream ORF start site) was overlapped by a peak (Fig. 1B). We chose to use as the reference coordinate the translation initiation codon rather than the transcription start site (TSS) because the latter has not been accurately determined for all genes. Then, our arbitrary definition of the promoter as the -350 to -100 region relative to the ATG codon was based on the distribution of the TSS-ATG distance for genes with an annotated TSS (median and average distance = 58 and 107 bp, respectively; see Supp. Fig. 1A).

Strikingly, Dal80 was found to bind to 1269 gene promoters (Fig. 1C-D and Supp. Table 3a). Figure 1E shows a snapshot of ChIP-Seq signals at the promoter of the *FUR4* gene, identified as bound by Dal80. The number of Dal80-bound promoters, corresponding to 22% of all protein-coding genes promoters, is much higher than expected given the hundred target genes that are generally cited for the GATA transcriptional activators (Godard et al., 2007; Scherens et al., 2006), that are presumed to bind to the same sites as Dal80. However, we noted that some peaks (221) overlapped with several promoters (471), mainly of divergent genes (442). In such cases, it is therefore possible that only one of the two divergent promoters is indeed targeted by Dal80, so that the number of Dal80-bound promoters could actually be a little less important than 1269.

As expected, genes related to previous nitrogen regulation screens were present within our list (Airoldi et al., 2016; Boer et al., 2003, 2007; Cox et al., 1999; Daugherty et al., 1993; Godard et al., 2007; Scherens et al., 2006; Tai et al., 2005; Wu et al., 2004)(Supp. Table 3b). Among the genes

showing Dal80-binding to the promoter, we also noticed a strong enrichment for cytoplasmic translation genes, as well as genes involved in small molecule biosyntheses, including amino acids (Supp. Table 3c).

Surprisingly, analysis of GATA site occurrence over Dal80-bound and unbound promoters revealed no major difference between the two classes (Fig. 1F), 48.2% and 51.3% of Dal80-bound and unbound promoters containing at least two GATA sites, respectively (see Supp. Table 3d). Intriguingly, 20% of Dal80-bound promoters do not contain any GATA site (Fig. 1F and Supp. Table 3d), indicating that Dal80 recruitment can also occur independently of the presence of consensus GATA sites (see Supp. Fig. 1B for visualization of Dal80 binding to a GATA-less promoter).

In summary, our ChIP-Seq analysis revealed that Dal80 binds to a set of promoters larger than previously expected, targeting biosynthetic functions and protein synthesis in addition to nitrogen catabolite repression.

Dal80 binds the intragenic region of a subset of genes

Metagene analysis revealed that the genes bound by Dal80 at the promoter also display a signal along the gene body, although this intragenic signal remains globally lower than in the promoter-proximal region (Fig. 1D). This observation prompted us to investigate the possibility that Dal80 also occupies the gene body, at least for a subset of genes.

We identified 189 genes showing Dal80 intragenic occupancy, according to a >75% overlap of the ORF by a Dal80-Myc[13] peak (Fig. 2A-B). Among them, 144 (76%) were also bound at the promoter (Fig. 2B). Conversely, only 11% of the Dal80-bound promoters (144/1269) showed downstream intragenic binding (Fig. 2B). On the other hand, 45 genes showing Dal80 intragenic binding were not bound at the promoter (Fig. 2B). Hence, we distinguished four classes of genes (Supp. Table 4): (*i*) those bound by Dal80 at the promoter only ("P" class; Fig. 2C), (*ii*) those showing both promoter and intragenic binding ("P & O" class; Fig. 2D), (*iii)* those bound across the ORF only ("O" class; Fig. 2E),

(*iv*) the unbound genes (Fig. 2F). Interestingly, we noted that the global Dal80-Myc[13] signal at the promoter was higher for the "P & O" class in comparison to the "P" class (compare Fig. 2C-D).

In conclusion, a subset of genes displays intragenic Dal80 binding, and in most cases, this correlates with strong Dal80 recruitment at the promoter.

Dal80 binding across gene bodies correlates with Dal80-sensitive, high expression levels

In order to correlate Dal80 binding with gene expression and Dal80-sensitivity in similar culture conditions, we performed total RNA-seq analysis in wild type and *dal80*Δ cells grown in medium with proline as the nitrogen source.

The expression of 546 genes was significantly altered in proline-grown *dal80*Δ cells compared to wild type (Fig. 3A, Supp. Table 5a), among which 232 and 314 genes were negatively and positively regulated by Dal80, respectively. The first group was enriched for genes involved in small molecule catabolic processes (Supp. Table 5b), while the positively regulated genes were mostly involved in amino acid biosynthetis (Supp. Table 5c). Again, we noticed an overlap between Dal80-repressed genes and nitrogen regulated genes that were identified in other genome-wide screens (Supp. Table 3b).

Analysis of mRNA levels in proline-grown cells revealed a significant correlation between Dal80 binding and regulation at the mRNA level: 39.6% (216/546) of Dal80-regulated genes were bound by Dal80 at their promoter and/or across the ORF, which is significantly more than expected if Dal80 binding and regulation were independent (P<0.00001, Chi-square test; Supp. Fig. 2A). More importantly, additional differences are emerging between the different classes of genes. First, gene expression was found to be globally similar between the unbound genes and those of the "O" class (P=0.15, Wilcoxon rank-sum test; Fig. 3B). Second, Dal80 binding at the promoter of genes correlated with significantly increased expression levels, in comparison to the unbound genes ($P < 2.2e^{-16}$, Wilcoxon rank-sum test; Fig. 3B). Third, the genes bound by Dal80 both at the promoter and across the ORF ("P & O" class) were even more expressed than the genes showing promoter binding only (P

= $1.3e^{-14}$, Wilcoxon rank-sum test; Fig. 3B). Together with the observation that genes of the "P & O" class globally showed higher Dal80-Myc^{13} ChIP-Seq signal at the promoter than those of the "P" class (Fig. 2C-D), our results indicate that Dal80 spreading across gene bodies correlates with a strong recruitment at the promoter, higher sensitivity to Dal80 and high expression in proline-containing medium. Consistently, metagene analysis of Dal80-regulated genes is very similar to that of "P & O" class (Supp. Fig. 2C).

In conclusion, our genome-wide analyses revealed a significant correlation between Dal80 recruitment to the chromatin and Dal80-sensitive expression. We also showed that Dal80 binding both at the promoter and across the gene body correlates with high levels of gene expression.

Dal80 binding across the ORF of well-characterized NCR-sensitive genes

In order to validate our genome-wide observations at the level of known NCR-sensitive genes, we characterized the binding profiles of Dal80-Myc^{13} along four NCR-sensitive genes of the "P & O" class (see snapshots of ChIP-Seq signals in Fig. 4A-D). We chose the NAD^+-dependent glutamate dehydrogenase gene (*GDH2*) and three permease-coding genes (*GAP1*, *DAL5* and *MEP2*), aware that their regulation could differ (Georis et al., 2011; Godard et al., 2007; Scherens et al., 2006) and that GATA sites are differently represented throughout these genes (Fig. 4A-D).

As expected from our genome-wide observations, Dal80 bound not only the promoter, but also the coding region of *DAL5*, *GAP1*, *GDH2* and *MEP2* in proline-grown cells (Fig. 4E and Supp. Fig. 3A). Binding of Dal80 to the *GDH2* ORF was reduced (Fig. 4E), which is consistent with the ChIP-Seq profile (Fig. 4D). We ruled out any erroneous interpretation of our data upon showing that, using the same experimental design, the Gal4-Myc^{13} protein only occupied the *GAL10* promoter, and not the coding region (Supp. Fig. 3B). Moreover, intragenic GATA factor binding did not rely on an intermediate RNA molecule, since the Dal80-Myc^{13} signal across the *MEP2* gene body was unaffected upon RNAse treatment of the chromatin extracts before the immunoprecipitation (Fig. 4F and Supp. Fig. 3C).

The results of the ChIP-qPCR show that the level of Dal80 binding also varied from one gene to another. As striking example, Dal80 was more present at the *GAP1* and *MEP2* ORFs, which are more highly expressed than *DAL5* in proline-grown cells (Supp. Table 5a). This is consistent with the correlation observed at the genomic level between Dal80 binding both to the promoter and across gene bodies and high expression levels (Fig. 3B). Altogether, these observations prompted an important mechanistic question: how can Dal80 be localized to gene bodies ?

Dal80 binding within gene bodies occurs independently of intragenic GATA sites

Like most NCR-sensitive genes, the four genes described above are all characterized by stretches of GATA sequences in their promoters (Fig. 4A-D). A possible explanation for Dal80 binding across gene bodies could be that functional, intragenic GATA sequences, directly recruit Dal80. Given the low complexity of the GATA consensus, the probability that a gene is devoid of intragenic GATA site is very low. Indeed, only 10% of all yeast ORFs have no GATA site, and most genes contain an average of 5 intragenic GATA sites (Supp. Fig. 3c). However, according to current models, GATA sites should be present in stretches of at least two in order to be functional *in vivo*, which is not the case within *DAL5*, *GDH2*, *GAP1* and *MEP2* ORFs, that contain some GATA sites, but relatively distant from each other. Interestingly, 23 of the 144 P&-O-bound genes did not contain any GATA site in their ORF, hence showing that Dal80 spreading across the body of genes does not necessarly require intragenic GATA sites, at least for these genes.

In order to test if the presence of a NCR-sensitive promoter could confer intragenic Dal80 binding across the body of a non-NCR-sensitive gene, devoid of any GATA cluster, we placed the *URA3* ORF under the control of the *MEP2* promoter, at the *MEP2 locus* (P_{MEP2}-*URA3*; Fig. 5A). Only two GATA sites are present in the *URA3* ORF, and separated by 300bp. In this strain, the expression of *URA3* became NCR-sensitive and followed wild type *MEP2* expression (Supp. Fig. 4A), correlating with PolII recruitment over the *URA3* ORF (Supp. Fig. 4B). In these conditions, we observed Dal80-Myc[13] binding at the promoter of *MEP2* and also across *URA3* (Fig. 5B). Importantly, Dal80 was not

detected across *URA3* when it was expressed from its native *locus*, under the control of its promoter (Fig. 5B). These results show that significant recruitment of Dal80 at an ORF region must be preceded by promoter recruitment.

Finally, we analyzed Dal80 binding at the *UGA4 locus*, another well-characterized NCR-sensitive gene, which is devoid of intragenic GATA site clusters (Fig. 5C). *UGA4* expression is induced by GABA (γ-aminobutyric acid) and is strongly repressed by Dal80 in the absence of the inducer (André et al., 1993). Therefore, to derepress *UGA4* without inducer and to analyze Dal80 binding along *UGA4*, we deleted the repressive C-terminal leucine zipper domain of Dal80 (Dal80ΔLZ-Myc^{13}), known for its homodimerization and heterodimerization with the Gzf3 transcriptional repressor (Coornaert et al., 1992; Svetlov and Cooper, 1998)(Fig. 5D). As a control experiment, *UGA4* expression was quantified in Dal80-Myc^{13} and Dal80ΔLZ-Myc^{13} cells. As expected, *UGA4* mRNA level was very low in Dal80-Myc^{13} cells, whereas deletion of Dal80 leucine zipper led to a strong derepression of *UGA4* (Supp. Fig. 4C). Analysis of PolII occupancy across *UGA4* using ChIP confirmed that elevated mRNA levels correlate with increased transcription (Supp. Fig. 4D). In these conditions, full-length Dal80-Myc^{13} binding was restricted to the *UGA4* promoter, while Dal80ΔLZ-Myc^{13} binding was detected at the promoter and across the body of *UGA4* (Fig. 5E), correlating with the elevated transcript and PolII levels. Surprisingly, the leucine zipper of Dal80 and consequently, its dimerization, needed for *UGA4* repression, were not required for its spreading across *UGA4* gene body. Importantly, these results show that promoter binding is not sufficient to confer intragenic binding, but suggest that transcription activation is required.

Together, these data indicate that Dal80 binding at the promoter and across the body of NCR-sensitive genes correlated with expression levels and required transcription activation.

Transcription elongation is necessary for Dal80 binding across MEP2

We hypothesized that the elongating PolII complex could be responsible for Dal80 spreading beyond Dal80-bound promoters.

To assess whether the transcriptionally active PolII complex is necessary for Dal80 binding across the *MEP2* gene, we used an *rpb1-1* strain, known to rapidly involve a cessation of PolII activity when cells are shifted to a restrictive temperature of 37°C (Kim et al., 2010; Nonet et al., 1987). We analyzed Dal80-Myc^{13} binding along *MEP2* in WT and *rpb1-1* cells. When *rpb1-1* cells were shifted at 37°C for 1hr, *MEP2* mRNA and PolII levels showed a 2-fold (Supp. Fig. 5A) and >10-fold decrease (Supp. Fig. 5B), respectively, reflecting the expected elongation defect when *rpb1-1* cells are shifted in non-permissive conditions. In the same conditions, we observed a significant >5-fold reduction of Dal80-Myc^{13} levels across the *MEP2* ORF, while the binding at the promoter was not affected (Fig. 6A-B). This result suggests that Dal80 spreading across the body of NCR-sensitive genes would be a consequence of the PolII-driven elongation process.

To get insights into the mechanism by which Dal80 associates to gene bodies, we tested whether it physically interacts with the transcriptionnaly engaged form of PolII (Fig. 6C). Total protein extracts from Dal80-Myc^{13} cells were immunoprecipitated with antibodies directed against the PolII CTD and its phospho-forms Ser2P and Ser5P, respectively characteristic of elongating and initiating PolII forms. All three antibodies enabled effective immunoprecipitation, whereas no antibody and nonspecific antibody controls generated a lower or no signal at all. Thus, Dal80 physically interacts with phosphoforms of the PolIII, suggesting a strong association with PolII engaged in active transcription from initiating to elongating polymerase.

Together, our data indicate that Dal80 spreading across the body of NCR-sensitive genes depends on active transcription and that Dal80 interacts with the transcriptioanlly active form of PolII, supporting a model where Dal80 spreading across the body of highly expressed, NCR-sensitive genes might be the result of Dal80-PolII association at post-initiation transcription phases.

Discussion

Eukaryotic GATA factors belong to a small but important family of DNA binding proteins involved in development and response to environmental changes in multicellular organisms in unicellular organisms, respectively. In yeast, four GATA factors are involved in Nitrogen Catabolite Repression (NCR), controlling gene expression in response to nitrogen source availability. One of them, the Dal80 repressor, itself NCR-sensitive, acts to modulate the intensity of NCR responses.

Over the past decade, a number of studies have screened the genome aiming at gathering an inventory of genes regulated by the nitrogen source. Although >500 genes have been shown to be differentially expressed upon change of the nitrogen source (Airoldi et al., 2016; Godard et al., 2007), the list of NCR-sensitive genes was reduced to about 100, based on their sensitivity to GATA factors (Boer et al., 2003, 2007; Godard et al., 2007; Scherens et al., 2006), suggesting that the number of Dal80 targets would be situated in that range. Here, our ChIP-Seq analysis identified more than 1200 Dal80-bound promoters, which considerably extends the list of potential Dal80 targets. In fact, the number of Dal80-bound promoters could even have been more important. Indeed, the GATA consensus binding site is rather simple and short, so that in yeast, a total number of 10,000 putative binding sites can be found in all protein-coding gene promoters, 2930 promoters having at least 2 GATA sites, which is thought to be a prerequisite for *in vivo* binding and function of the GATA factors. The difference between the number of promoters with ≥2 GATA sites and the number of Dal80-bound promoters suggests the existence of a selectivity for Dal80 recruitment. This selectivity could rely on promoter architecture and/or chromatin structure, conditioning the requirement for auxiliary DNA binding factors that would stabilize Dal80 at some promoters. Moreover, although we observed a significant correlation between Dal80 binding and regulation, the expression of most of the Dal80-bound genes was not affected in a *dal80Δ* mutant strain. Again, Dal80-dependence for transcribing these genes, as well as their NCR sensitivity, could require the presence of cofactors which are not produced or inactive under the tested growth conditions. In mammals, GATA factors also display an extraordinary complexity in the relationships between binding and expression regulation. Like Dal80,

GATA-1 and GATA-2 only occupy a small subset of their abundant binding motif throughout the genome, and the presence of the conserved binding site is insufficient to cause GATA-dependent regulation in most instances (Hasegawa et al., 2016). GATA-1 binding kinetics, stoichiometry and heterogenous complex formations, conditioned by composite promoter architecture, influence its transcriptional activity and hence diversify gene expression profiles (Hasegawa et al., 2016).

Given the high conservation at the amino acid level between the DNA binding domain of the four yeast NCR GATA factors, it is likely that they all recognize identical sequences (GATAA, GATAAG or GATTAG). This consensus has been largely validated in the past using gene reporter experiments, mutational analyses and *in vitro* binding experiments on naked DNA. Nonetheless, of the 1269 bound promoters, 48% contained at least two GATA sites, a proportion that is not different from that observed among unbound promoters, and the amount of GATA sites per promoter was not different between the two groups either. In addition, Dal80 recruitment was found to occur independently of the presence of GATA sites in 20% of Dal80-bound promoters, as also previously observed in mammalian cells (Cheng et al., 2009; Fujiwara et al., 2009). Future experiments will be required to decipher the molecular mechanism by which Dal80 can be recruited to these GATA-less promoters.

Unexpectedly, although Dal80 has always been described as a repressor, we identified 314 genes that are positively regulated by Dal80 (their expression is significantly decreased upon Dal80 deletion). These genes are distinct from the NCR-sensitive genes : only 4 belong to the NCR-sensitive list of (Godard et al., 2007) and are significantly enriched in amino acid biosynthetic processes, resembling the amino acid starvation response mediated by the Gcn4 transcriptional activator. Interestingly, the promoter of 122/314 Dal80-activated genes contain Gcn4-binding sites (Supp. Fig. 3c), and this group of 314 Dal80-activated genes is significantly enriched for genes regulated by the General Amino Acid Control (GAAC; YeastMine Gene List, Publication Enrichment, $p<1.6e-13$), through the Gcn4 activator. Interconnections between NCR and GAAC have already been demonstrated, mostly at the level of nitrogen catabolism control: 1-a large number of non-preferential nitrogen sources leads to increased transcription of GAAC targets (Godard et al., 2007);

and 2- Gcn4 contributes, with Gln3, to the expression of NCR-sensitive genes (Staschke et al., 2010; Valenzuela et al., 2001). However, this is the first time that evidence are provided suggesting a positive function for the negative GATA factor at the level biosynthetic gene expression.

The most striking and unexpected finding of this work is the observation that for 189 genes, Dal80 occupied the gene body ("O" and "P & O" classes). The 45 genes displaying Dal80 intragenic occupancy with no recruitment at the promoter ("O" class) were globally neither Dal80-sensitive nor NCR-sensitive, similar to the unbound genes. Furthermore, we also noted that a substantial fraction of these genes correspond to small dubious ORFs, near or even overlapping the promoter, Dal80-bound, of the adjacent gene. In these cases, the limited resolution of the ChIP-Seq technique, combined to the small size of these genes, might have allowed them to pass through the filters of the procedure to identify the genes showing Dal80 intragenic binding. Overall, these observations suggest that the existence of the "O" class is likely to be physiologically irrelevant. Dal80 binding at the promoter and spreading across the body of the 144 genes belonging to the "P & O" class correlated with high expression levels and sensitivity to Dal80-dependent repression. These observations were experimentally confirmed using ChIP experiments, at the level of well-characterized NCR-sensitive genes. Our data indicate that Dal80 spreading across genes bodies does not rely on intragenic GATA sites. Rather, promoter binding appears to be required but not sufficient: Dal80 spreading across genes depends on active transcription. Indeed, the observation that impaired PolII elongation correlates with decreased intragenic binding further indicates that Dal80 binding across gene bodies depends on ongoing transcription elongation. Consistently, we detected a physical interaction between Dal80 and transcriptionnaly active forms of PolII. Together, our data leads us to propose a model where Dal80 could travel from the promoter of highly expressed, NCR-sensitive genes through the gene body by accompanying the elongating PolII complex (Figure 7). Additional investigations will be required to define which domain of Dal80 is responsible for the interaction with the transcription machinery, to determine whether there is any causal relationship

between Dal80 intragenic binding and high expression levels, and to decipher the potential role of Dal80 during transcription elongation.

Whereas the binding of elongation factors across gene bodies has been thoroughly documented (Sims et al., 2004), it has also been described for few specific transcription factors. For example, Gal4 was reported to bind to its consensus DNA target within the *ACC1* ORF, but the authors concluded that the observed transcriptional repression of the *ACC1* gene was most likely resulting from random *GAL4* binding "noise" over the genome, thus having no physiological explanation for this ORF-bound transcription factor (Li and Johnston, 2001). Likewise, Gcn4 was detected across the *PHO8* ORF, with concomitant recruitment of the SAGA complex, but without any impact on gene expression (Topalidou and Thireos, 2003).

A recent study has identified CTD phosphorylation of polII as a hub that optimizes transcriptome changes to adequately balance optimal growth and stress tolerance responses (Chasman et al., 2014). The addition of nitrogen to nitrogen-limited cells rapidly results in the transient overproduction of transcripts required for protein translation (stimulated growth) whereas accelerated mRNA degradation favours rapid clearing of the most abundant transcripts, like those involved in high affinity permease production, that are highly expressed NCR-sensitive genes, for example (Airoldi et al., 2016). The involvement of the Nrd1-Nab3-Sen1 (NNS) and TRAMP complexes in these regulatory responses has been envisioned very recently (Bresson et al., 2017; Merran and Corden, 2017). Thus, a possible role of Dal80 (and possibly of the other GATA factors) binding along highly expressed genes could be to transmit nutritional signals to the RNA decay machinery for rapid and efficient remodeling of gene expression in response to changing nitrogen supplies. Here, an attractive model (Figure 7) would hypothesize that Dal80 binding to polII could prevent the decay of the dedicated stress-response mRNAs to enhance their cellular life-cycle by regulating/repressing RNA decay factors that might associate otherwise on the polymerase. Other elongation-related processes, like histone modification and chromatin remodeling (Henikoff and Shilatifard, 2011; Tessarz and Kouzarides, 2014), or mRNA export (Meinel et al., 2013), can also be proposed.

In this respect, in human cells, GATA factors are also reported to occupy non-canonical sites within the genome, further reinforcing that they can be recruited to the chromatin independently of their motif (Cheng et al., 2009; Fujiwara et al., 2009). In addition, 43% of the GATA1 peaks were collected among exon, introns and 3'UTR of coding genes in human erythroleukemia cells (Fujiwara et al., 2009). It is tempting to hypothesize that GATA factors could have a dual or synergistic role during transcription, i.e. recruiting/stabilizing the PIC complex as for any classical transcription factor in the promoter/enhancer regions and promoting elongation at a post initiation step interacting with the RNAPII. Interestingly, GATA1 in erythropoid cells interacts with the mediator (Stumpf et al., 2006), a complex also known to travel with PolII (Allen and Taatjes, 2015).

Author contributions

AR, IG and ED conceived and designed the study. AR constructed strains and performed experiments. AR, MW and AM designed NGS experiments and data analysis. MD and CG performed bioinformatics analyses. IG, MW, ED and AM wrote the article. IG and ED supervised the project.

Acknowledgments

We are grateful to all members of our labs for discussions and critical reading of the manuscript. We thank F. Vierendeels for excellent technical assistance. This work has benefited from the facilities and expertise of the NGS platform of Institut Curie, supported by the Agence Nationale de la Recherche (ANR-10-EQPX-03, ANR10-INBS-09-08) and the Canceropôle Ile-de-France. A. Morillon's lab is supported by the Agence Nationale de la Recherche (DNA-Life) and the European Research Council (EpincRNA starting grant, DARK consolidator grant). Work by IG and ED is funded by the CoCoF and FNRS (FRFC 2.4547.11). Work by AR was funded by a FRIA fellowship.

Conflict of interest

The authors declare that they have no competing interests.

References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Andre, B., Hein, C., Grenson, M., and Jauniaux, J.C. (1993). Cloning and expression of the UGA4 gene coding for the inducible GABA-specific transport protein of Saccharomyces cerevisiae. Mol Gen Genet *237*, 17-25.

Andre, B., Talibi, D., Soussi Boudekou, S., Hein, C., Vissers, S., and Coornaert, D. (1995). Two mutually exclusive regulatory systems inhibit UASGATA, a cluster of 5'-GAT(A/T)A-3' upstream from the UGA4 gene of Saccharomyces cerevisiae. Nucleic acids research *23*, 558-564.

Bertram, P.G., Choi, J.H., Carvalho, J., Ai, W., Zeng, C., Chan, T.F., and Zheng, X.F. (2000). Tripartite regulation of Gln3p by TOR, Ure2p, and phosphatases. The Journal of biological chemistry *275*, 35727-35733.

Bresnick, E.H., Katsumura, K.R., Lee, H.Y., Johnson, K.D., and Perkins, A.S. (2012). Master regulatory GATA transcription factors: mechanistic principles and emerging links to hematologic malignancies. Nucleic acids research *40*, 5819-5831.

Bresnick, E.H., Lee, H.Y., Fujiwara, T., Johnson, K.D., and Keles, S. (2010). GATA switches as developmental drivers. The Journal of biological chemistry *285*, 31087-31093.

Bysani, N., Daugherty, J.R., and Cooper, T.G. (1991). Saturation mutagenesis of the UASNTR (GATAA) responsible for nitrogen catabolite repression-sensitive transcriptional activation of the allantoin pathway genes in Saccharomyces cerevisiae. Journal of bacteriology *173*, 4977-4982.

Cheng, Y., Wu, W., Kumar, S.A., Yu, D., Deng, W., Tripic, T., King, D.C., Chen, K.B., Zhang, Y., Drautz, D.*, et al.* (2009). Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. Genome research *19*, 2172-2184.

Coffman, J.A., Rai, R., Cunningham, T., Svetlov, V., and Cooper, T.G. (1996). Gat1p, a GATA family protein whose production is sensitive to nitrogen catabolite repression, participates in transcriptional activation of nitrogen-catabolic genes in Saccharomyces cerevisiae. Molecular and cellular biology *16*, 847-858.

Coffman, J.A., Rai, R., Loprete, D.M., Cunningham, T., Svetlov, V., and Cooper, T.G. (1997). Cross regulation of four GATA factors that control nitrogen catabolic gene expression in Saccharomyces cerevisiae. Journal of bacteriology *179*, 3416-3429.

Cooper, T.G., Rai, R., and Yoo, H.S. (1989). Requirement of upstream activation sequences for nitrogen catabolite repression of the allantoin system genes in Saccharomyces cerevisiae. Molecular and cellular biology *9*, 5440-5444.

Coornaert, D., Vissers, S., Andre, B., and Grenson, M. (1992). The UGA43 negative regulatory gene of Saccharomyces cerevisiae contains both a GATA-1 type zinc finger and a putative leucine zipper. Curr Genet *21*, 301-307.

Cunningham, T.S., and Cooper, T.G. (1991). Expression of the DAL80 gene, whose product is homologous to the GATA factors and is a negative regulator of multiple nitrogen catabolic genes in Saccharomyces cerevisiae, is sensitive to nitrogen catabolite repression. Molecular and cellular biology *11*, 6205-6215.

Cunningham, T.S., and Cooper, T.G. (1993). The Saccharomyces cerevisiae DAL80 repressor protein binds to multiple copies of GATAA-containing sequences (URSGATA). Journal of bacteriology *175*, 5851-5861.

Cunningham, T.S., Dorrington, R.A., and Cooper, T.G. (1994). The UGA4 UASNTR site required for GLN3-dependent transcriptional activation also mediates DAL80-responsive regulation and DAL80 protein binding in Saccharomyces cerevisiae. Journal of bacteriology *176*, 4718-4725.

Cunningham, T.S., Svetlov, V.V., Rai, R., Smart, W., and Cooper, T.G. (1996). Gln3p is capable of binding to UAS(NTR) elements and activating transcription in Saccharomyces cerevisiae. Journal of bacteriology *178*, 3470-3479.

Daugherty, J.R., Rai, R., el Berry, H.M., and Cooper, T.G. (1993). Regulatory circuit for responses of nitrogen catabolic gene expression to the GLN3 and DAL80 proteins and nitrogen catabolite repression in Saccharomyces cerevisiae. Journal of bacteriology *175*, 64-73.

Descrimes, M., Ben Zouari, Y., Wery, M., Legendre, R., Gautheret, D., and Morillon, A. (2015). VING: a software for visualization of deep sequencing signals. BMC research notes *8*, 419.

Georis, I., Feller, A., Vierendeels, F., and Dubois, E. (2009). The yeast GATA factor Gat1 occupies a central position in nitrogen catabolite repression-sensitive gene activation. Molecular and cellular biology *29*, 3803-3815.

Georis, I., Tate, J.J., Cooper, T.G., and Dubois, E. (2011). Nitrogen-responsive regulation of GATA protein family activators Gln3 and Gat1 occurs by two distinct pathways, one inhibited by rapamycin and the other by methionine sulfoximine. The Journal of biological chemistry *286*, 44897-44912.

Georis, I., Tate, J.J., Vierendeels, F., Cooper, T.G., and Dubois, E. (2015). Premature termination of GAT1 transcription explains paradoxical negative correlation between nitrogen-responsive mRNA, but constitutive low-level protein production. RNA biology *12*, 824-837.

Godard, P., Urrestarazu, A., Vissers, S., Kontos, K., Bontempi, G., van Helden, J., and Andre, B. (2007). Effect of 21 different nitrogen sources on global gene expression in the yeast Saccharomyces cerevisiae. Molecular and cellular biology *27*, 3065-3086.

Grass, J.A., Boyer, M.E., Pal, S., Wu, J., Weiss, M.J., and Bresnick, E.H. (2003). GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. Proceedings of the National Academy of Sciences of the United States of America *100*, 8811-8816.

Hahn, S., and Young, E.T. (2011). Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics *189*, 705-736.

Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.B., Reynolds, D.B., Yoo, J.*, et al.* (2004). Transcriptional regulatory code of a eukaryotic genome. Nature *431*, 99-104.

Horak, C.E., Mahajan, M.C., Luscombe, N.M., Gerstein, M., Weissman, S.M., and Snyder, M. (2002). GATA-1 binding sites mapped in the beta-globin *locus* by using mammalian chIp-chip analysis. Proceedings of the National Academy of Sciences of the United States of America *99*, 2924-2929.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Kim, T.S., Liu, C.S., Yassour, M., Holik, J., Friedman, N., Buratowski, S., and Rando, O.J. (2010). RNA polymerase mapping during stress responses reveals widespread nonproductive transcription in yeast. Genome biology *11*, R75.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods *9*, 357-359.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I.*, et al.* (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science (New York, N.Y *298*, 799-804.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature *424*, 147-151.

Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953-961.

Lowry, J.A., and Atchley, W.R. (2000). Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. Journal of molecular evolution *50*, 103-115.

MacIsaac, K.D., Wang, T., Gordon, D.B., Gifford, D.K., Stormo, G.D., and Fraenkel, E. (2006). An improved map of conserved regulatory sites for Saccharomyces cerevisiae. BMC bioinformatics *7*, 113.

Magasanik, B., and Kaiser, C.A. (2002). Nitrogen regulation in Saccharomyces cerevisiae. Gene *290*, 1-18.

Medina-Rivera, A., Defrance, M., Sand, O., Herrmann, C., Castro-Mondragon, J.A., Delerce, J., Jaeger, S., Blanchet, C., Vincens, P., Caron, C.*, et al.* (2015). RSAT 2015: Regulatory Sequence Analysis Tools. Nucleic acids research *43*, W50-56.

Minehart, P.L., and Magasanik, B. (1991). Sequence and expression of GLN3, a positive nitrogen regulatory gene of Saccharomyces cerevisiae encoding a protein with a putative zinc finger DNA-binding domain. Molecular and cellular biology *11*, 6216-6228.

Nonet, M., Scafe, C., Sexton, J., and Young, R. (1987). Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. Molecular and cellular biology *7*, 1602-1611.

Omichinski, J.G., Clore, G.M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S.J., and Gronenborn, A.M. (1993). NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. Science (New York, N.Y *261*, 438-446.

Rai, R., Genbauffe, F.S., Sumrada, R.A., and Cooper, T.G. (1989). Identification of sequences responsible for transcriptional activation of the allantoate permease gene in Saccharomyces cerevisiae. Molecular and cellular biology *9*, 602-608.

Rowen, D.W., Esiobu, N., and Magasanik, B. (1997). Role of GATA factor Nil2p in nitrogen regulation of gene expression in Saccharomyces cerevisiae. Journal of bacteriology *179*, 3761-3766.

Scherens, B., Feller, A., Vierendeels, F., Messenguy, F., and Dubois, E. (2006). Identification of direct and indirect targets of the Gln3 and Gat1 activators by transcriptional profiling in response to nitrogen availability at short and long term. FEMS Yeast Res *6*, 777-791.

Soussi-Boudekou, S., Vissers, S., Urrestarazu, A., Jauniaux, J.C., and Andre, B. (1997). Gzf3p, a fourth GATA factor involved in nitrogen-regulated transcription in Saccharomyces cerevisiae. Mol Microbiol *23*, 1157-1168.

Stanbrough, M., and Magasanik, B. (1996). Two transcription factors, Gln3p and Nil1p, use the same GATAAG sites to activate the expression of GAP1 of Saccharomyces cerevisiae. Journal of bacteriology *178*, 2465-2468.

Starich, M.R., Wikstrom, M., Arst, H.N., Jr., Clore, G.M., and Gronenborn, A.M. (1998). The solution structure of a fungal AREA protein-DNA complex: an alternative binding mode for the basic carboxyl tail of GATA factors. Journal of molecular biology *277*, 605-620.

Svetlov, V.V., and Cooper, T.G. (1998). The Saccharomyces cerevisiae GATA factors Dal80p and Deh1p can form homo- and heterodimeric complexes. Journal of bacteriology *180*, 5682-5688.

ter Schure, E.G., van Riel, N.A., and Verrips, C.T. (2000). The role of ammonia metabolism in nitrogen catabolite repression in Saccharomyces cerevisiae. FEMS Microbiol Rev *24*, 67-83.

Tripic, T., Deng, W., Cheng, Y., Zhang, Y., Vakoc, C.R., Gregory, G.D., Hardison, R.C., and Blobel, G.A. (2009). SCL and associated proteins distinguish active from repressive GATA transcription factor complexes. Blood *113*, 2191-2201.

van Helden, J., Andre, B., and Collado-Vides, J. (1998). Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. Journal of molecular biology *281*, 827-842.

Venters, B.J., and Pugh, B.F. (2009). A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. Genome research *19*, 360-371.

Wery, M., Descrimes, M., Vogt, N., Dallongeville, A.S., Gautheret, D., and Morillon, A. (2016). Nonsense-Mediated Decay Restricts LncRNA Levels in Yeast Unless Blocked by Double-Stranded RNA Structure. Molecular cell *61*, 379-392.

Zambelli, F., Pesole, G., and Pavesi, G. (2013). Motif discovery and transcription factor binding sites before and after the next-generation sequencing era. Briefings in bioinformatics *14*, 225-237.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Main Figure Titles and Legends

Figure 1. Genome-wide identification of Dal80-bound promoters.

A. Overview of the ChIP-Seq analysis. Biological duplicates of FV78 (Dal80-Myc^{13}) and 25T0b (no tag) cells were grown to mid-log phase in proline-containing medium, and then harvested. After chromatin extraction and sonication, Dal80-Myc^{13} was immunoprecipitated using α-Myc antibody. Co-precipitated DNA fragments were purified and used to construct ChIP-Seq libraries. After sequencing of the libraries, signals were computed using uniquely mapped reads. Dal80-bound regions were identified using a peak-calling procedure using MACS2.

B. Identification of Dal80-bound promoters. After peak-calling, Dal80-bound promoters were identified on the basis of a >75% overlap of the -100 to -350 regions (relative to the downstream ORF start site) by the peak (represented in as a red dashed line).

C. Heatmap view of the ChIP-Seq signal in the ATG +/- 500 bp region for the 1269 genes identified as bound by Dal80 at the promoter, in untagged (left) and Dal80-Myc^{13} cells (right). The colour goes from yellow to red as the signal increases.

D. Metagene view of the ChIP-Seq signal along the ATG +/- 600 bp region for the 1269 genes identified as bound by Dal80 at the promoter (solid lines) and for the unbound genes (dashed lines), in untagged (black) and Dal80-Myc^{13} cells (blue). For each group of genes, normalized coverage (tag/nt) for each gene was piled up, and the average signal was plotted. The shading surrounding each line denotes the 95% confidence interval.

E. Snapshot of ChIP-Seq signals along the *FUR4 locus*. Densities (tag/nt) are shown for the untagged (black line) and Dal80-Myc^{13} (blue line) strains. ORFs and the Dal80-Myc^{13} ChIP-Seq peak are represented by grey arrows and the red dashed line, respectively. The position (and orientation) of each GATA site is represented by vertical segments above (sense GATA sites) or below (antisense GATA sites) the *locus* line. The snapshot was produced using the VING software (Descrimes *et al.*, 2015).

F. Number of GATA sites in the promoter of Dal80-unbound (4522) and promoter-bound (1268) genes. The analysis was performed using RSAT (Medina-Rivera *et al.*, 2015), across the -500 to -1 region (relative to the ATG codon of the downstream ORF).

Figure 2. Dal80 binds the intragenic region of a subset of genes.

A. Identification of genes showing Dal80 intragenic binding.

B. Venn diagram showing the number of genes that are bound by Dal80 at the promoter and/or across the ORF.

C. Metagene view of the ChIP-Seq signal along the ATG +/- 600 bp region for the 1125 genes identified as bound by Dal80 at the promoter only ("P"), in untagged (black) and Dal80-Myc[13] cells (blue). MetaSignal computation was as described in Fig. 1D. The shading surrounding each line denotes the 95% confidence interval.

D. Same as above for the 144 genes bound by Dal80 at the promoter and across the ORF ("P & O").

E. Same as above for the 45 genes bound by Dal80 across the ORF only ("O").

F. Same as above for the 4484 genes that are not bound by Dal80.

Figure 3. Dal80 binding across gene bodies correlates with Dal80-sensitive, high expression levels.

A. Scatter plot of densities (tag/nt, $\log_2$ scale) for genes in WT and *dal80Δ* cells grown in proline-containing medium. Total RNA was extracted from exponentially growing biological replicates of 25T0b (WT) and FV080 (*dal80Δ*). After rRNA depletion, strand-specific RNA-Seq libraries were constructed and then sequenced. Tag densities were computed using uniquely mapped reads. Dal80-regulated genes were identified on the basis of a min. 2-fold change (up or down) in the *dal80Δ* mutant relative to the WT control, with a P <0.05 upon differential expression analysis using DESeq (Anders and Huber, 2010). Unaffected (n=5252), Dal80-repressed (n=232) and Dal80-activated (n=314) genes are shown as grey, red and blue dots, respectively.

B. Box-plot of densities (tag/nt, log_2 scale) in the WT strain grown in proline-containing medium, for genes of the unbound, "O", "P" and "P & O" classes. *P*-values obtained upon Wilcoxon rank-sum test are indicated.

Figure 4. Dal80 binding across the ORF of well-characterized NCR-sensitive genes.

A-D. Position of GATA sites (vertical segments above -sense GATA sites- or below -antisense GATA sites- the *locus* line) and qPCR probes (black and grey lettered segments) along 4 NCR-sensitive loci: *DAL5* (A), *GAP1* (B), *GDH2* (C) and *MEP2* (D).

E. Cells of FV078 (wt *DAL80-MYC*[13]) strain were grown in proline as unique nitrogen source. Anti-Myc ChIP was performed as described in Experimental procedures. Each histogram represents the average (x10 000) of the value IP/IN (input or total chromatin) of at least 3 independent cultures on which two IPs were performed. The associated error bars correspond to the standard error. Black histograms represent qPCR performed on promoter regions whereas grey ones indicate qPCR data obtained within genes. Dotted line represents the highest no tag control levels. Primers are described in Supp. Table 2: a, $DAL5_{P1-P2}$; b, $DAL5_{O3-O4}$; c, $GAP1_{P1-P2}$; d, $GAP1_{O9-O10}$; e, $GDH2_{P1-P2}$; f, $GDH2_{O5-O6}$; g, $MEP2_{P9-P10}$; h, $MEP2_{O1-O2}$.

F. ChIP analysis was conducted as in Panel E, with or without RNase treatment before immunoprecipitation.

Figure 5. Dal80 binding within gene bodies occurs independently of intragenic GATA sites.

A. Schematic representation of the P_{MEP2}-*URA3 locus*, GATA sites (vertical segments above -sense GATA sites- or below -antisense GATA sites- the *locus* line) and qPCR probe positions (black segments, *MEP2* promoter; grey segments, *URA3* ORF).

B. Occupancy of Dal80-Myc[13] at the *MEP2* promoter and *URA3* ORF at the *MEP2* and *URA3 loci* in wild type and P_{MEP2}-*URA3* strains. Cells of *DAL80-MYC*[13] wild type (FV078) and P_{MEP2}-*URA3* (FV808) strains were cultivated at 29°C in the presence of proline (Pro) as unique nitrogen source. ChIP

analysis was conducted as in Fig. 4. The first histogram summarizes the data obtained by qPCR probing the promoter of *MEP2* (a, $MEP2_{P9-P10}$), whereas the two following histograms are generated by qPCR performed with oligos positioned in the ORF of *URA3* (b, $URA3_{O1-O2}$; c, $URA3_{O3-O4}$), either at the *URA3 locus* (in wild type strain), or both at the *URA3* and *MEP2 loci*. Dotted line represents the highest no tag control levels.

C. Schematic representation of the *UGA4 locus*, GATA sites (vertical segments above -sense GATA sites- or below -antisense GATA sites- the *locus* line) and qPCR probe positions (black segments, promoter; grey segments, ORF).

D. Wild type and truncated (ΔLZ) versions of Dal80-Myc[13] used for the ChIP experiment at the *UGA4 locus*.

E. Occupancy of the *UGA4 locus* by Dal80-Myc[13]. Cells of FV078 (*DAL80-MYC[13]*; wt) and FV136 (*DAL80ΔLZ-MYC[13]*; ΔLZ) strains were grown at 29°C in the presence of proline (Pro) as unique nitrogen source. ChIP analysis was conducted as described in Fig. 4. Black histograms represent qPCR performed with promoter-specific primers (a, $UGA4_{P1-P2}$; b, $UGA4_{P9-P10}$) whereas grey ones indicate qPCR data obtained with ORF-specific primers (c, $UGA4_{O1-O2}$; d, $UGA4_{O3-O4}$) at the *UGA4 locus*.

Figure 6. Transcription elongation is necessary for Dal80 binding across *MEP2*.

A. Schematic representation of the *MEP2 locus*, GATA sites (vertical segments above -sense GATA sites- or below -antisense GATA sites- the *locus* line), and qPCR probe positions (black segments, promoter; grey segments, ORF).

B. Occupancy of Dal80-Myc[13] at the *MEP2 locus* when transcription elongation is impaired. Wild type (FV673) or *rpb1-1* (FV675) *DAL80-MYC[13]* cells were grown at 29°C in the presence of proline (Pro) as unique nitrogen source, and shifted at 37°C for one hour at mid-log phase. ChIP analysis was conducted as in Fig. 4. The first 2 black columns represent qPCR with 2 primer pairs within the *MEP2* promoter (a, $MEP2_{P5-P6}$; b, $MEP2_{P13-P14}$) and the 2 following grey ones correspond to qPCR data obtained with 2 primer pairs within the *MEP2* ORF (c, $MEP2_{O11-O12}$ d, $MEP2_{O9-O10}$).

C. Coimmunoprecipitation of Dal80-Myc[13] with different phosphoforms of PolII. Total proteins were extracted, immunoprecipitated with the indicated antibodies, and subjected to anti-Myc western blot analysis.

Figure 7. Model summarizing the correlations observed between Dal80 binding, Dal80 dependency and expression levels.

STAR Methods

Contact for Reagent and Resource Sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Isabelle Georis (igeoris@ulb.ac.be). Bioinformatics and genome wide dataset requests could also be addressed to antonin.morillon@curie.fr for rapid processing.

Experimental Model and Subject Details

Experiments were conducted using *S. cerevisiae* strains of the FY genetic background. The strains used are listed in Supp. Table 1. Dal80 was tagged with 13 copies of the *c-myc* epitope (Myc[13]) as described (Longtine et al., 1998) using primers listed in Supp. Tables 1 and 2. The P_{MEP2}-*URA3* allele in strains FV806-808 was created by amplification of the *URA3* gene using the same strategy, with primers listed in Supp. Tables 1 and 2.

Cultures were grown at 29°C to mid-log phase (A_{660nm} = 0.5) in YNB (without amino acids or ammonia) minimal medium containing the indicated nitrogen source at a 0.1% final concentration, glucose (3%) and the appropriate supplements (20 µg/ml uracil, histidine and tryptophan) to cover auxotrophic requirements.

Method Details

Chromatin immunoprecipitation

Cell extracts and immunoprecipitations were conducted as described (Georis et al., 2009) using primers listed in Supp. Table 2. The cells (100-ml cultures grown to an absorbance (A660 nm = 0.6) corresponding to 6 × 106 cells/ml) were treated with 1% formaldehyde for 30 min at 25 °C and mixed by orbital shaking. Glycine was then added to a final concentration of 500 mm and incubation continued for 5 min. The cells were collected, washed once with cold 10 mm Tris-HCl, pH 8, washed once with cold FA-SDS buffer (50 mm HEPES-KOH, pH 7.5, 150 mm NaCl, 1 mm EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 1 mm phenylmethylsulfonyl fluoride), and resuspended in

1 ml of cold FA-SDS buffer. An equal volume of glass beads (0.5 mm in diameter) was added, and the cells were disrupted by vortexing for 30 min in a cold room. The lysate was diluted into 4 ml of FA-SDS buffer, and the glass beads were discarded. The cross-linked chromatin was then pelleted by centrifugation (17,000 × g for 35 min), washed for 60 min with FA-SDS buffer, resuspended in 1.6 ml of FA-SDS buffer for 15 min at 4 °C, and sonicated three times for 30 s. each (Bioruptor, Diagenode) to yield an average DNA fragment size of 700 base pairs. Finally, the sample was clarified by centrifugation at 14,000 × g for 30 min and diluted 4-fold in FA-SDS buffer, and aliquots of the resultant chromatin containing solution were stored at –80 °C. Myc[13]-tagged proteins were immunoprecipitated by incubating 100 µl of the chromatin containing solution for 180 min at 4°C with 2 µl of mouse anti-Myc antibodies (SC-40, Santa Cruz) prebound to 10 µl of Dynabeads Pan Mouse IgG (Dynal) according to the manufacturer's instructions. Immune complexes were washed six times in FA-SDS buffer and recovered by treating with 50 µl of Pronase Buffer (25 mm Tris, pH 7.5, 5 mm EDTA, 0.5% SDS) at 65 °C with agitation. Input (IN) and immunoprecipitated (IP) fractions were then subjected to Pronase treatment (0.5 mg/ml; Roche Applied Science) for 60 min at 37°C, and formaldehyde cross-links were reversed by incubating the eluates overnight at 65°C. Finally, the samples were treated with RNase (50 µg/ml) for 60 min at 37°C. DNA from the IP fractions was purified using the High Pure PCR Product Purification Kit (Roche Applied Science) and eluted in 50 µl of 20 mm Tris buffer, pH 8. IN fractions were boiled 10 min and diluted 500-fold with no further purification prior to quantitative PCR analysis.

Quantitative RT-PCR

Quantitative RT-PCR was performed as described previously (Georis et al., 2009) using primers listed in Supp. Table 2. Total RNA was extracted from 4-ml cultures and cDNA was generated from 100 to 500 ng of total RNA using a RevertAid H Minus first-strand cDNA synthesis kit with oligo$(dT)_{18}$ primers from Fermentas using the manufacturer's recommended protocol. cDNAs were subsequently quantified by RT-PCR using the Maxima SYBR green qPCR master mix from Fermentas.

Co-Immunoprecipitation

Cultures (100 ml) were harvested, washed once in 50 mM Tris, pH 8, and resuspended in 1ml of buffer (50 mM Tris, pH 8, 150 mM NaCl, 5 mM EDTA, 0.05% NP-40, 1 mM phenylmetyhlsulfonyl fluoride, and complete protease inhibitor cocktail tablets [Roche]). Lysis was performed by shaking with 425-600 µm acid-washed glass beads (Sigma) on an IKA Vibrax VXR orbital shaker at maximum speed for 30 min at 4°C. Cell debris and glass beads were removed by centrifugation. Immunoprecipitation was performed by incubating 200 µl of total cell extracts with 20 µl of Dynabeads PAN mouse immunoglobulin G (Invitrogen) that were preinucbated with anti-HA (SCBT, SC-7392), anti-CTD (SCBT, CTD4H8), anti-Ser2P (BioLegend, H5) or anti-Ser5P (BioLegend, H14) antibodies and 20 µl of 1% phosphate-buffered saline-bovine serum albumin for 2 h under orbital shaking (800 rpm) at 30°C. Immune complexes were washed three times in lysis buffer, eluted by boiling in sodium dodecyl sulfate (SDS) sample buffer, and loaded on SDS-polyacrylamide gel for anti-Myc Western blotting.

ChIP-Seq analysis and Peak-calling

ChIP-Seq analysis was performed from two biological replicates of proline-grown 25T0b (no tag) and FV078 (Dal80-Myc[13]) cells. For each condition, libraries were prepared from 10 ng of “input” or “IP” DNA using the TruSeq ChIP Sample Preparation Kit (Illumina). Single-read sequencing (50 nt) of the libraries was performed on a HiSeq 2500 sequencer.

Reads were uniquely mapped to the *S. cerevisiae* S288C reference genome using Bowtie2 v2.1.0 (Langmead and Salzberg, 2012), with a tolerance of 1 mismatch in seed alignment. Tags densities were normalized on the total number of uniquely reads mapped.

Dal80-bound regions were identified through a peak-calling procedure using version 2.0.9 of MACS (Zhang et al., 2008), with a minimum false discovery rate (FDR) of 0.001.

Total RNA-Seq

For each strain and condition, total RNA was extracted from two biological replicates using standard hot phenol procedure, ethanol-precipitated, resuspended in nuclease-free H_2O (Ambion) and quantified using a NanoDrop 2000c spectrophotometer. Ribosomal RNAs were depleted from 1 µg of total RNA using the RiboMinus™ Eukaryote v2 Kit (Life Technologies). After concentration using the Ribominus™ Concentration Module (Life Technologies), rRNA-depleted RNA was quantified using the Qubit RNA HS Assay kit (Life Technologies). In parallel, rRNA depletion efficiency and integrity of both total and rRNA-depleted RNA were checked by analysis in a RNA 6000 Pico chip, in a 2100 bioanalyzer (Agilent). Strand-specific total RNA-Seq libraries were prepared from 125 ng of rRNA-depleted RNA using the TruSeq® Stranded Total RNA Sample Preparation Kit (Illumina), following manufacturer's instructions. Paired-end sequencing (2 x 50 nt) of the libraries was performed on a HiSeq 2500 sequencer. Sequenced reads were mapped to the reference genome using version 2.0.6 of TopHat (Kim et al., 2013), as described (Wery et al., 2016). Tags densities were normalized on the total number of reads uniquely mapped on ORFs. Differential expression analysis was performed using DESeq (Anders and Huber, 2010).

Quantification and Statistical Analysis

Statistical details are to be found in the figure legends. Error bars correspond to standard error. Statistical significance tests were carried out using the Student's t test (Microsoft Excel) when indicated.

Data and Software Availability

Sequence data can be accessed at the NCBI Gene Expression Omnibus using accession numbers GSE86307 and GSE86325. Genome browsers for visualization of processed ChIP-Seq and RNA-Seq data are accessible at http://vm-gb.curie.fr/dal80.

Supplemental Item Titles and Legends

Supplemental Table 1. Strains used in this study.

Supplemental Table 2. Primers used in this study.

Supplemental Table 3 - Related to Figure 1. Genome-wide identification of Dal80-bound promoters. *Sheet a.* List of gene promoters bound by Dal80. *Sheet b.* Overlap of genes identified in our screens and in previous genome-wide expression screens. *Sheet c.* GO term analysis of Dal80-bound promoters. *Sheet d.* Whole genome GATA site contents of promoters and ORFs, correlated with Dal80 binding, regulation, and NCR status and presence of Gcn4-binding site.

Supplemental Table 4 - Related to Figure 2. Dal80 binds the intragenic region of a subset of genes. Inventory of gene promoters and ORFs bound by Dal80 and their intersection.

Supplemental Table 5 - Related to Figure 3. Dal80 binding across gene bodies correlates with Dal80-sensitive, high expression levels. *Sheet a.* List of Dal80-regulated genes on proline. *Sheet b.* Gene ontology term analysis of the 232 genes repressed by Dal80 on proline. *Sheet c.* Gene ontology term analysis of the 314 genes activated by Dal80 on proline.

Supplemental Figure 1 - Related to Figure 1. Genome-wide identification of Dal80-bound promoters.

A. Box-plot of the distance between the annotated TSS and ORF start site (translation initiation codon, ATG) for protein-coding genes.

B. Snapshot of ChIP-Seq signals along a GATA-less *locus* (*ALD6*). Densities (tag/nt) are shown for the untagged (black line) and Dal80-Myc[13] (blue line) strains. ORFs and the Dal80-Myc[13] ChIP-Seq peak

are represented by grey arrows and the red dashed line, respectively. The snapshot was produced using the VING software (Descrimes et al., 2015).

Supplemental Figure 2 – Related to Figure 3. Dal80 binding across gene bodies correlates with Dal80-sensitive, high expression levels.

A. Contingency table showing the number of Dal80-regulated or unaffected genes (defined upon RNA-Seq analysis in WT and *dal80Δ* cells) among the Dal80-bound and unbound genes (identified upon ChIP-Seq analysis in untagged and Dal80-Myc[13] cells). Results that were experimentally observed or that were expected in case of independence are indicated in bold and in brackets, respectively.

B. Metagene view of the ChIP-Seq signal along the 216 Dal80-regulated and -bound genes (solid lines), and the 330 Dal80-regulated and -unbound genes (dashed lines), in the untagged (black) and Dal80-Myc[13] (blue) strains. The shading surrounding each line denotes the 95% confidence interval.

Supplemental Figure 3 – Related to Figure 4. Dal80 binding across the ORF of well-characterized NCR-sensitive genes – Control experiments.

A. Cells of FV078 (wt *DAL80-MYC*[13]) strain were grown in glutamine (Gln) or proline (Pro) as unique nitrogen sources. Anti-Myc ChIP was performed as described in Experimental procedures. Each histogram represents the average (x10 000) of the value IP/IN (input or total chromatin) of at least 3 independent cultures on which two IPs were performed. The associated error bars correspond to the standard error. Black histograms represent qPCR performed on promoter regions whereas grey ones indicate qPCR data obtained within genes. Primers are described in Supp. Table 2: $DAL5_{P1-P2}$, $DAL5_{O3-O4}$, $GAP1_{P1-P2}$, $GAP1_{O9-O10}$, $GDH2_{P1-P2}$, $GDH2_{O5-O6}$, $MEP2_{P9-P10}$, $MEP2_{O1-O2}$.

B. Gal4-Myc[13] binding at the *GAL1-10* promoter. Schematic representation of the *GAL1-10 locus*. Position of the Gal4-binding sites (vertical segments across the *locus* line) and qPCR probes (black and grey lettered segments) are indicated. Cells of FV437 (wt *GAL4-MYC*[13]) strain were grown in YP

with glucose or galactose as unique carbon sources. Anti-Myc ChIP was performed as described in Experimental procedures. Each histogram represents the average (x10 000) of the value IP/IN (input or total chromatin) of at least 3 independent cultures on which two IPs were performed. The associated error bars correspond to the standard error. Black histograms represent qPCR performed on promoter regions whereas grey ones indicate qPCR data obtained within genes. Primers are described in Supp. Table 2.

C. Binding of Dal80-Myc[13] at the *MEP2 locus*, in untreated or RNase-treated cells. Schematic representation of the *MEP2 locus*. Position of the GATA sites (vertical segments above -sense GATA sites- or below -antisense GATA sites- the *locus* line) and qPCR probes (black and grey lettered segments) are indicated. Cells of untagged (25T0b) or FV078 (*DAL80-MYC*[13]; wt) strains were grown at 29°C in the presence of proline (Pro) as unique nitrogen source. ChIP analysis was conducted with or without RNase treatment before immunoprecipitation. The first 3 black columns represent qPCR with 3 primer pairs within the *MEP2* promoter $MEP2_{P5\text{-}P6}$; $MEP2_{P13\text{-}P14}$; $MEP2_{P9\text{-}P10}$) and the 3 following grey ones correspond to qPCR data obtained with 3 primer pairs within the *MEP2* ORF ($MEP2_{O1\text{-}O2}$; $MEP2_{O11\text{-}O12}$ $MEP2_{O9\text{-}O10}$).

Supplemental Figure 4 – Related to Figure 5. Dal80 binding within gene bodies occurs independently of intragenic GATA sites – Control experiments.

A. Expression of the non-NCR-sensitive gene *URA3* under the control of P_{MEP2}. Cells of wild type (25T0b) and P_{MEP2}-*URA3* (FV806) strains were cultivated at 29°C in the presence of proline (Pro) as unique nitrogen source. The expression of *URA3* is normalized to the expression of the *TBP1* gene. Each column corresponds to the mean value of data obtained for at least 2 independent cultures and the error bars correspond to the standard error.

B. PolII occupancy at the *URA3* and P_{MEP2}-*URA3 loci*. Cells of *DAL80-MYC*[13] wild type (FV078) and P_{MEP2}-*URA3* (FV808) strains were cultivated at 29°C in the presence of glutamine (Gln) or proline (Pro) as unique nitrogen sources. Anti-polII (CTD4H8) ChIP was performed as described in Experimental

procedures. Each histogram represents the average (x10 000) of the value IP/IN (input or total chromatin) of at least 3 independent cultures on which two IPs were performed. Histograms represent qPCR performed with $URA3_{O1-O2}$ primers.

C. *UGA4* expression in a *DAL80ΔLZ* mutant strain. Cells of FV078 (*DAL80-MYC*[13]; wt) and FV136 (*DAL80ΔLZ-MYC*[13]; ΔLZ) strains were cultivated at 29°C in the presence of proline (Pro) as unique nitrogen source. Total RNA was isolated and *TBP1*-normalized *UGA4* mRNA levels were quantified by qRT-PCR using $UGA4_{O1-O2}$ primers. Histograms represent the averages of at least 3 independent experiments and the associated error bars correspond to the standard error.

D. Occupancy of the *UGA4 locus* by PolII. Cells of FV078 (*DAL80-MYC*[13]; wt) and FV136 (*DAL80ΔLZ-MYC*[13]; ΔLZ) strains were grown at 29°C in the presence of proline (Pro) as unique nitrogen source. Anti-polII (CTD4H8) ChIP analysis was conducted as in Panel B using $UGA4_{O1-O2}$ primers.

Supplemental Figure 5 – Related to Figure 6. Transcription elongation is necessary for Dal80 binding across *MEP2* – Control experiments.

A. Effect of PolII elongation defects on *MEP2* expression. Wild type (FV673) or *rpb1-1* (FV675) *DAL80-MYC*[13] cells were grown at 29°C in the presence of glutamine (Gln) or proline (Pro) as unique nitrogen sources, and shifted at 37°C for one hour at mid-log phase. Total RNA was isolated and *TBP1*-normalized *MEP2* mRNA levels were quantified by qRT-PCR using $MEP2_{O9-O10}$ primers. Histograms represent the average of at least 3 independent experiments and the associated error bars correspond to the standard error.

B. PolII occupancy at the *MEP2 locus*. Wild type (FV673) or *rpb1-1* (FV675) *DAL80-MYC*[13] cells were grown at 29°C in the presence of glutamine (Gln) or proline (Pro) as unique nitrogen sources, and shifted at 37°C for one hour at mid-log phase. ChIP analysis was conducted as in Supp. Fig. 4B using $MEP2_{O9-O10}$ primers.

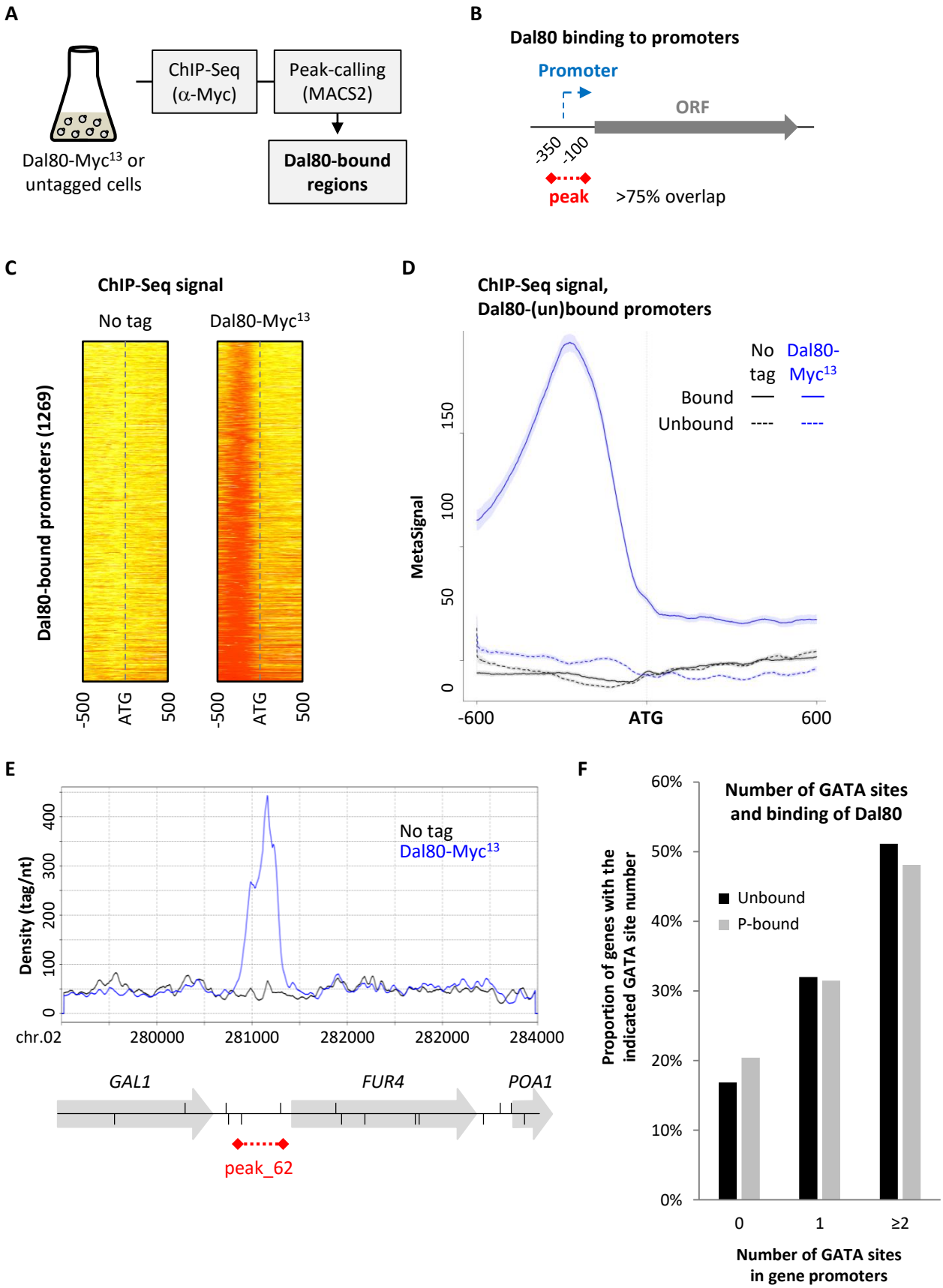

Figure 1. Genome-wide identification of Dal80-bound promoters.

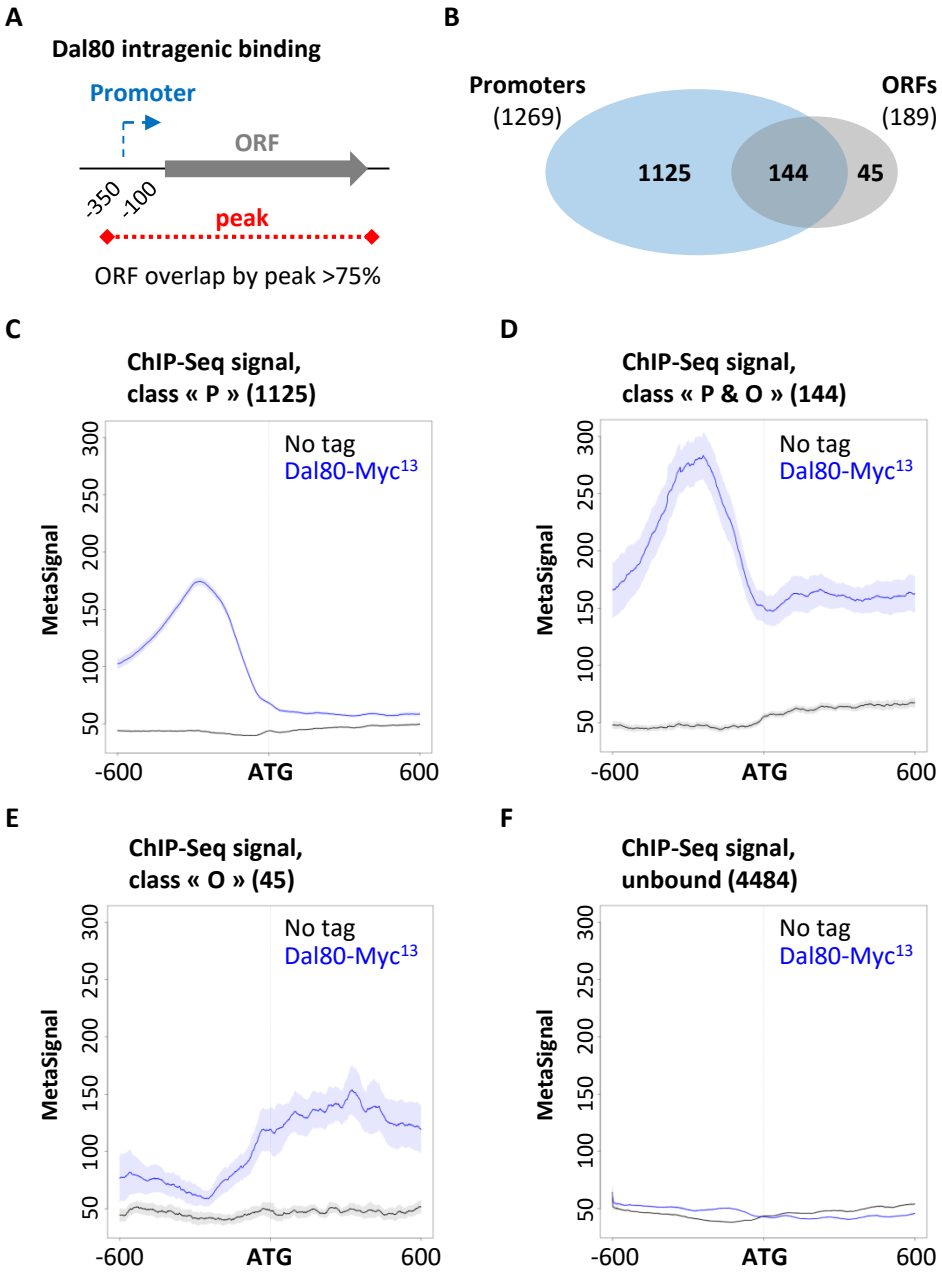

Figure 2. Dal80 binds the intragenic region of a subset of genes.

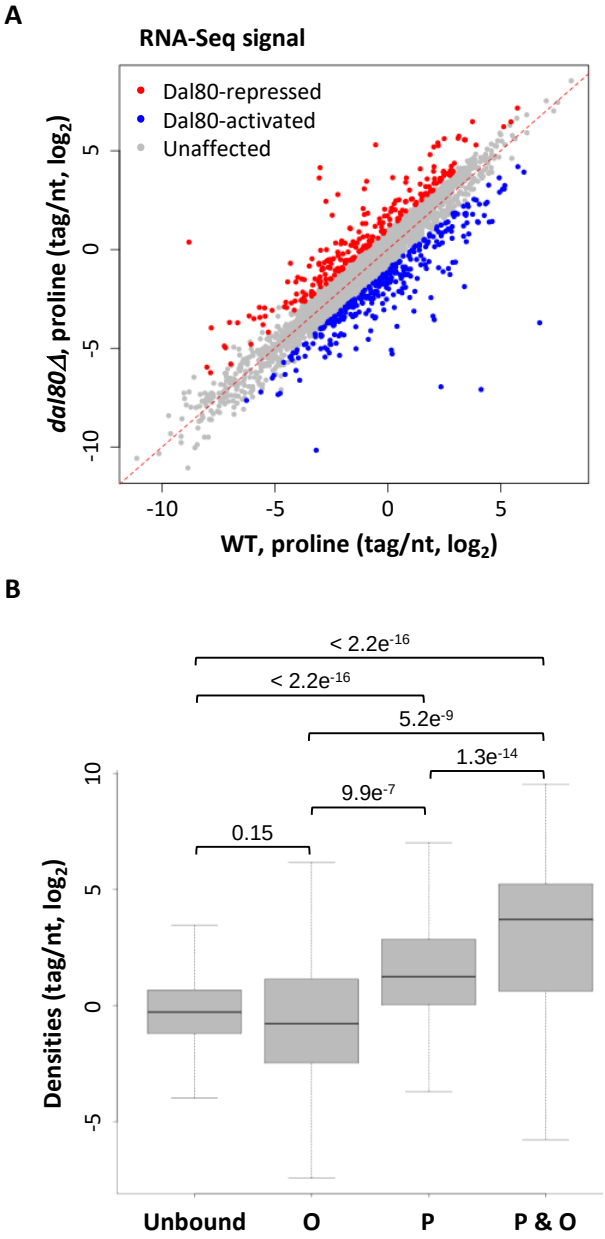

Figure 3. Dal80 binding across gene bodies correlates with Dal80-sensitive, high expression levels.

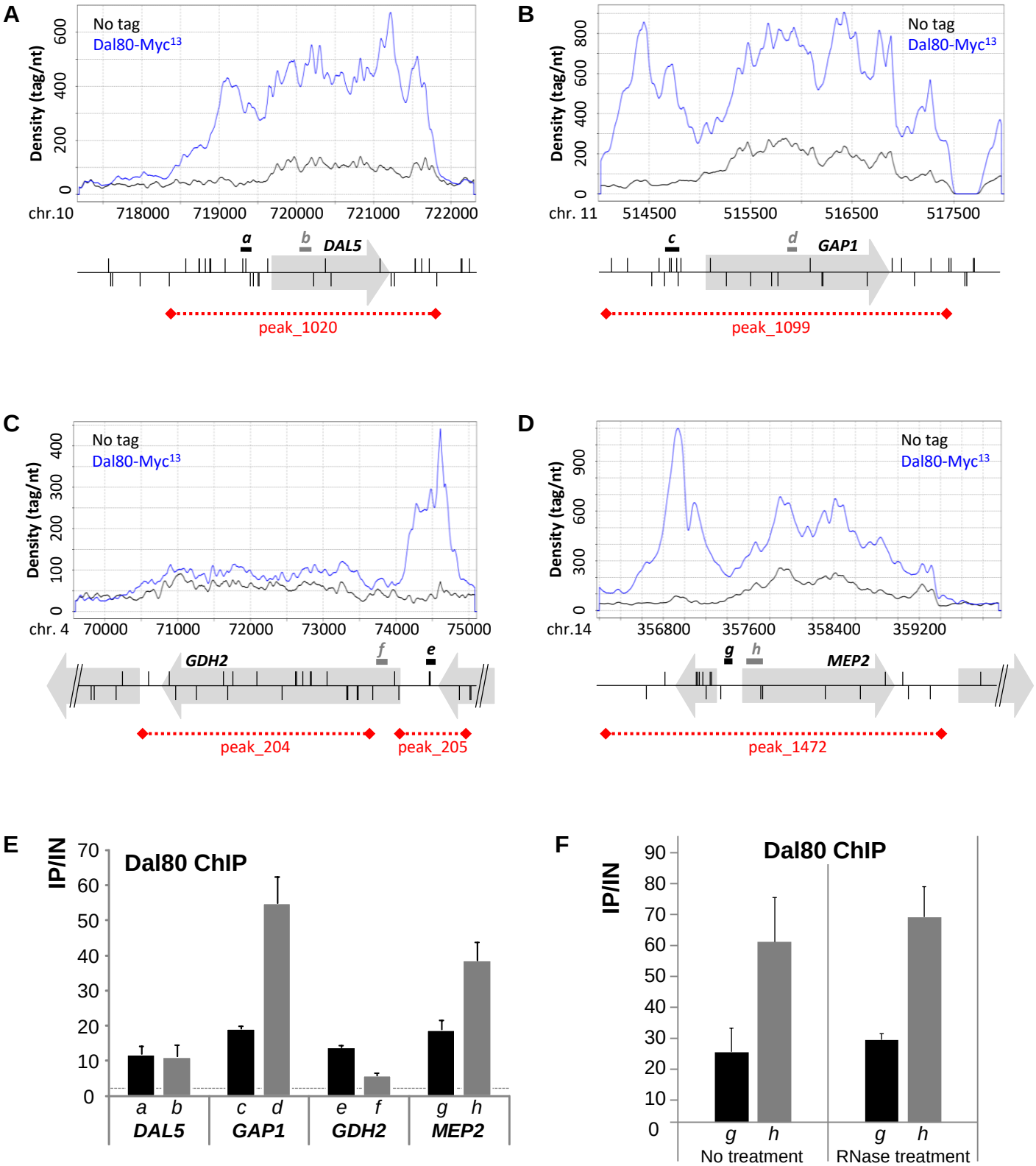

Figure 4. Dal80 binding across the ORF of well-characterized NCR-sensitive genes.

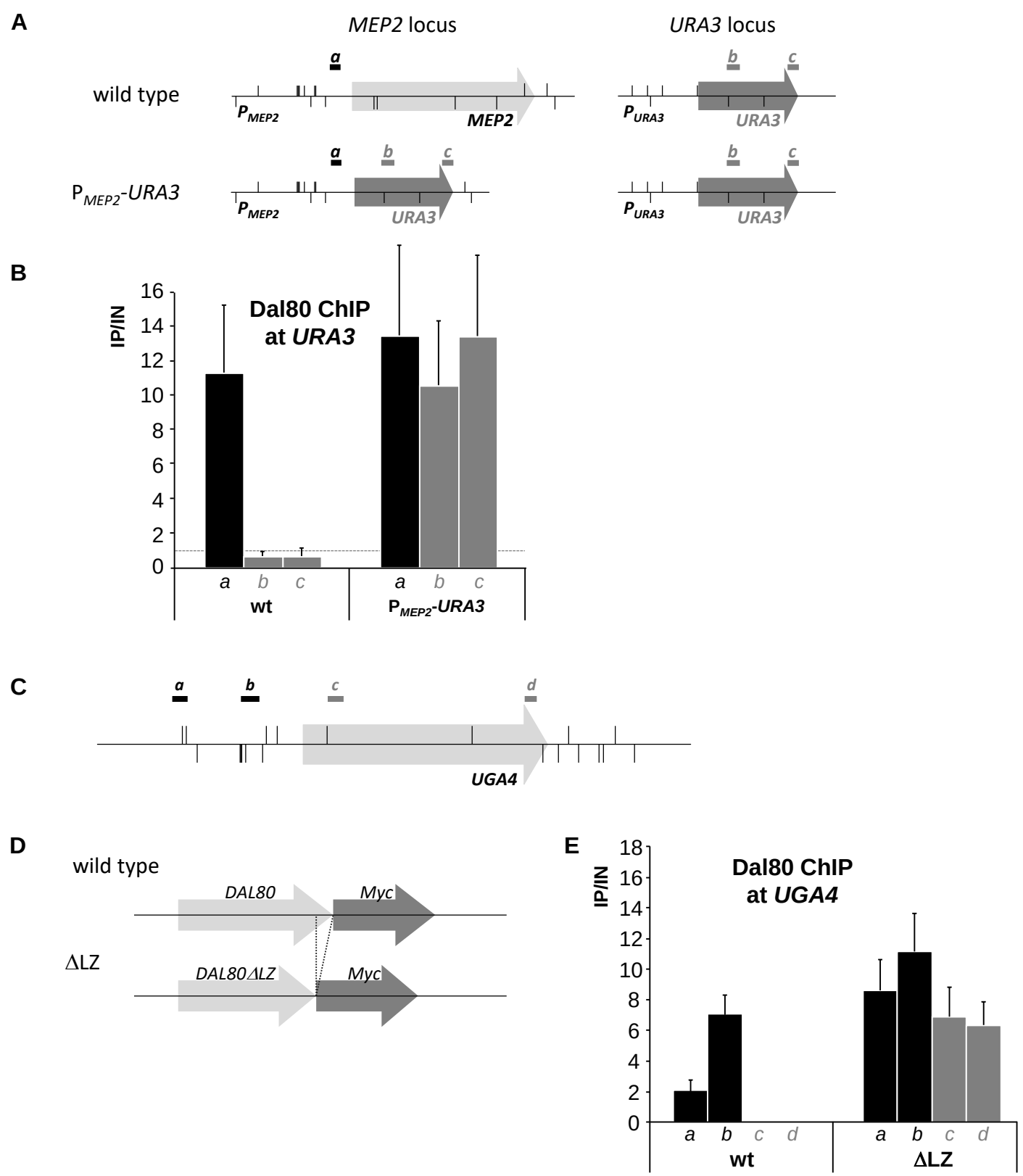

Figure 5. Dal80 binding within gene bodies occurs independently of intragenic GATA sites.

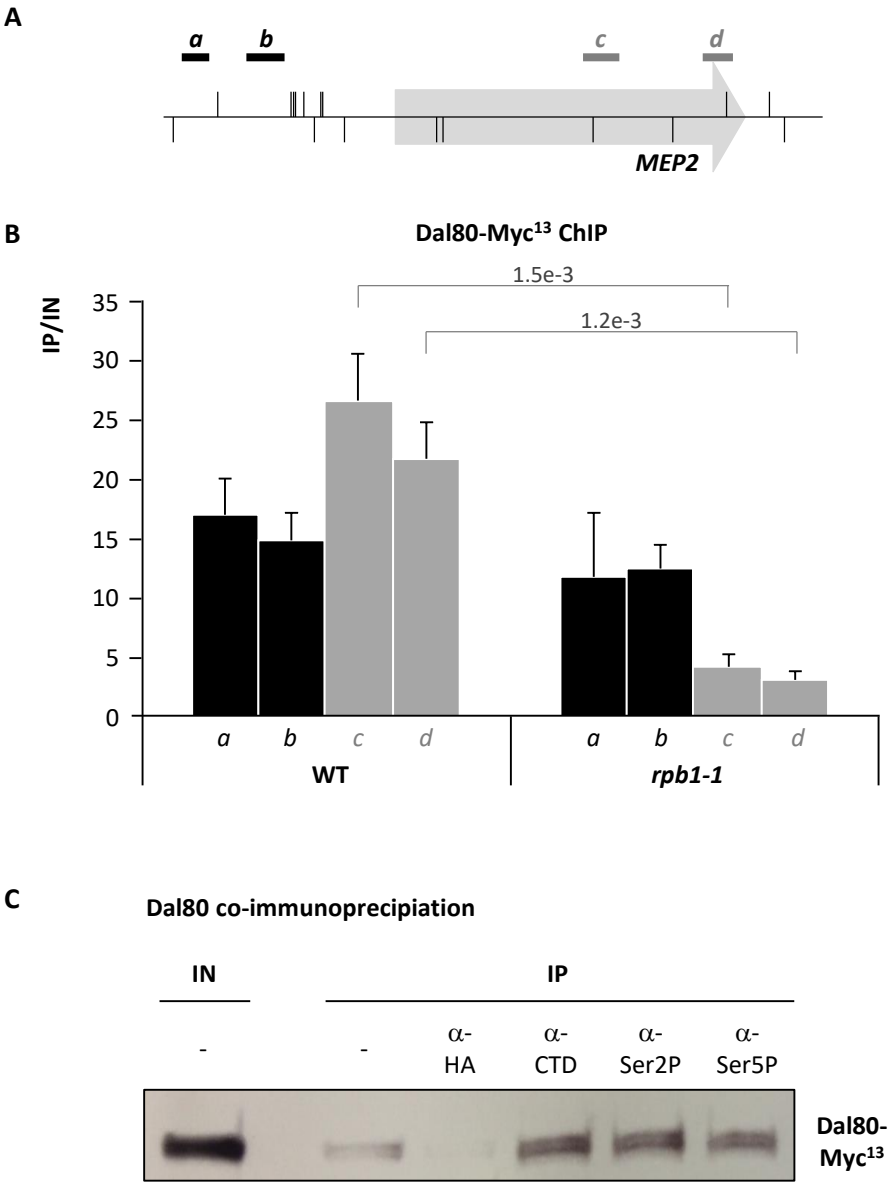

Figure 6. Transcription elongation is necessary for Dal80 binding across *MEP2*.

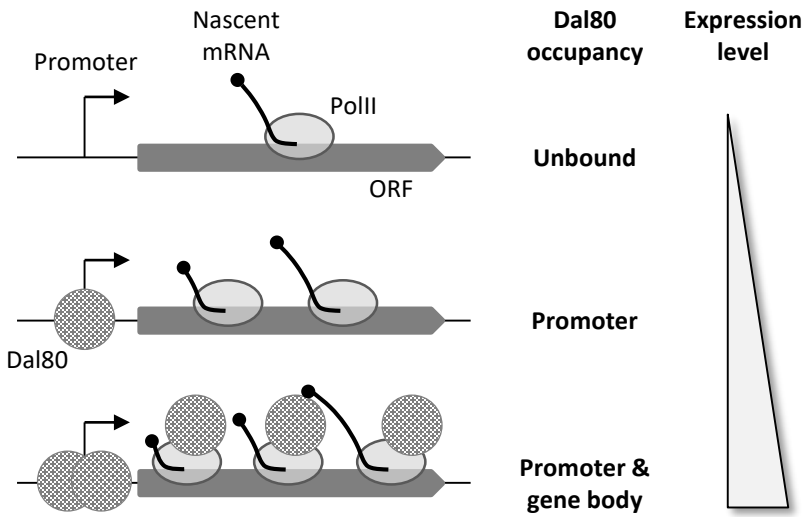

Figure 7. Model summarizing the correlations observed between Dal80 binding, Dal80 dependency and expression levels.