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 (lead contact), antonin.morillon@curie.fr

Abstract

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.

26 Our work suggests that GATA factors could play dual, synergistic roles during transcription initiation
27 and elongation, promoting efficient remodeling of the gene expression program in response to
28 environmental changes.
29

30 ## Introduction

31 In eukaryotes, gene transcription by RNA polymerase II (PolII) is initiated by the binding of specific
32 transcription factors to double stranded DNA. The yeast transcription factors target regulatory regions
33 called UAS or URS (for *U*pstream *A*ctivating/*R*epressing *S*equences), generally directly adjacent to the
34 core promoter. The generated regulatory signals converge at the core promoter where they permit
35 the regulation of PolII recruitment via the 'TATA box-binding protein' and associated general
36 transcription factors (Hahn & Young, 2011; Levine & Tjian, 2003). The transcription factor binding sites
37 are usually short sequences ranging from 8 to 20 bp (Zambelli, Pesole, & Pavesi, 2013). They are most
38 often similar but generally not identical, differing by some nucleotides from one another (Zambelli et
39 al., 2013), making it sometimes difficult to predict whether a given UAS will function as such *in vivo*.
40 GATA factors constitute a family of transcription factors highly conserved among eukaryotes and
41 characterized by the presence of a DNA binding domain which consists of four cysteines (fitting the
42 consensus sequence $CX_2CX_{17\text{-}18}CX_2C$) coordinating a zinc ion followed by a basic carboxy-terminal tail
43 (Omichinski et al., 1993). *In vitro*, this conserved binding domain allows the recognition of consensus
44 GATA DNA sequences (GATAA, GATAAG or GATTAG), found in the promoter of numerous genes, which
45 was demonstrated in various organisms either with direct or indirect methods (T. S. Cunningham,
46 Dorrington, & Cooper, 1994; Horak et al., 2002; Omichinski et al., 1993; Stanbrough & Magasanik,
47 1996; Starich, Wikström, Arst, Clore, & Gronenborn, 1998). Structure elucidation of GATA factor
48 binding domains in complex with their DNA target suggest an important role for the basic tail in
49 sequence specificity determination (Omichinski et al., 1993; Starich et al., 1998).
50 Since its discovery 40 years ago in chicken cells, the family of GATA factors was extended in human

51 cells and represents master regulators of hematopoiesis and cancer (Bresnick, Katsumura, Lee,
52 Johnson, & Perkins, 2012). However, although approximately 7 million GATA motifs can be found in
53 the human genome, the GATA factors occupy only 0.1–1% of them. Conversely, other regions are
54 occupied by GATA factors despite lacking the consensus motif (Cheng et al., 2009; Tripic et al., 2009).
55 In addition, GATA factors can swap among them for the same motif and switch from active or
56 repressive transcriptional activity. All these observations raised the main paradigm on understanding
57 the selective recruitment and activity of GATA factors and what the determinants of their residency
58 on the chromatin are (Bresnick, Lee, Fujiwara, Johnson, & Keles, 2010; Grass et al., 2003).
59 In yeast, the family of GATA transcription factors contains over 10 members (Lowry & Atchley, 2000).
60 Four of them are implicated in the regulation of *N*itrogen *C*atabolite *R*epression- (NCR-) sensitive
61 genes, the expression of which is repressed in the presence of a preferred nitrogen source (glutamine,
62 asparagine, ammonia) and derepressed when only poor nitrogen sources (e.g. proline, leucine, urea)
63 are available (Magasanik & Kaiser, 2002). These key GATA factors involved in NCR signaling are two
64 activators (Gln3 and Gat1/Nil1) and two repressors (Gzf3/Nil2/Deh1 and Dal80/Uga43) (J A Coffman,
65 Rai, Cunningham, Svetlov, & Cooper, 1996; Coornaert, Vissers, Andre, & Grenson, 1992; T S
66 Cunningham & Cooper, 1991; Minehart & Magasanik, 1991; Rowen, Esiobu, & Magasanik, 1997;
67 Soussi-Boudekou, Vissers, Urrestarazu, Jauniaux, & André, 1997). Moreover, the expression of *DAL80*
68 and *GAT1* is also NCR-sensitive, which implies cross- and autogenous regulations of the GATA factors
69 in the NCR mechanisms (Jonathan A. Coffman et al., 1997; I. Georis, Feller, Vierendeels, & Dubois,
70 2009; Isabelle Georis, Tate, Vierendeels, Cooper, & Dubois, 2015; Soussi-Boudekou et al., 1997). Under
71 nitrogen limitation, expression of *DAL80* is highly induced (T S Cunningham & Cooper, 1991), and Dal80
72 enters the nucleus where it competes with the two GATA activators for the same binding sites (André
73 et al., 1995; Jonathan A. Coffman et al., 1997; T. S. Cunningham et al., 1994). Although initially
74 described as being active under nitrogen abundance (Rowen et al., 1997; Soussi-Boudekou et al.,
75 1997), the Gzf3 repressor also localizes to NCR-sensitive promoters in conditions of activation (I. Georis
76 et al., 2009).

77 *In vitro*, the Gln3 and Gat1 activators bind, presumably as monomers, to single GATA sequences
78 (Thomas S. Cunningham, Svetlov, Rai, Smart, & Cooper, 1996), like their orthologous vertebrate
79 counterparts. On the other hand, the Dal80 repressor can form homo-and heterodimers with Gzf3,
80 thanks to the presence of an N-terminal leucine zipper (Svetlov & Cooper, 1998) and, consistently,
81 Dal80 binds *in vitro* to two GATA sequences (Jonathan A. Coffman et al., 1997; T. S. Cunningham &
82 Cooper, 1993; T. S. Cunningham et al., 1994; Svetlov & Cooper, 1998). *In vivo*, GATA factor binding site
83 recognition also appears to require repeats of GATA motifs within promoters. Indeed, synthetic
84 constructions containing a portion of the NCR-sensitive *DAL5* promoter have shown in a β-
85 galactosidase assay that at least two copies of the consensus are required to support efficient gene
86 activation. This led to the actual fuzzy definition of UAS_{NTR}, consisting in two GATA sites located
87 sufficiently close to one another to present a binding platform for GATA factors (Bysani, Daugherty, &
88 Cooper, 1991; Cooper, Rai, & Yoo, 1989; Rai, Genbauffe, Sumrada, & Cooper, 1989).
89 *In vivo*, GATA ChIP analyses have been carried out genome-wide in yeast (Harbison et al., 2004; Lee,
90 2002; MacIsaac et al., 2006; Venters & Pugh, 2009). They allowed gaining insights into the GATA factor
91 gene network through the identification of direct targets and characterization of the DNA elements
92 required to render a gene NCR-sensitive. However, these studies were not performed in activating
93 conditions, when all GATA factors are expressed, localized in the nucleus and active, so that the current
94 list of GATA factor targets are likely to be underestimated. On another hand, bioinformatic analyses
95 have shown that, since GATA sequences are short, they can be found almost everywhere throughout
96 the genome. Therefore, based on the sole criteria of the presence of repeated GATA sequences in
97 yeast promoters, a third of the yeast genes could hypothetically be NCR regulator targets (van Helden,
98 André, & Collado-Vides, 1998). Such GATA motif repetitions have been found in the promoter of 91
99 genes, inducible by GATA activators in absence of a good nitrogen source, supposed to be directly
100 targeted by the GATA activators (Scherens, Feller, Vierendeels, Messenguy, & Dubois, 2006).
101 Nevertheless, the functionality of these supposed UAS still needs to be directly demonstrated *in vivo*
102 (Hahn & Young, 2011).

103 Here, we provide the first genome-wide identification of Dal80 targets in yeast, in physiological
104 conditions where Dal80 is fully expressed and active. Using a ChIP-Seq approach combined to a
105 bioinformatic peak-calling procedure, we defined the exhaustive set of Dal80-bound promoters, which
106 turned out to be much larger than anticipated. Our data indicate that at some promoters, Dal80
107 recruitment occurs independently of GATA sites. Strikingly, Dal80 was also detected across the body
108 of a subset of genes bound at the promoter, globally correlating with high and Dal80-sensitive
109 expression. Mechanistic single-gene experiments confirmed the Dal80 binding profiles, further
110 indicating that Dal80 spreading across gene bodies occurs independently of intragenic GATA sites but
111 requires transcription elongation. Finally, co-immunoprecipitation experiments revealed that Dal80
112 physically interacts with both intitiating and elongating form of PolII. Our work suggests that in addition
113 to their well-established role at the (pre)inititation stage, GATA factors could be involved in
114 transcription elongation, promoting efficient remodeling of the gene expression program in response
115 to environmental changes.
116

117 ## Results

118 **Genome-wide identification of Dal80-bound promoters**

119 In order to determine the genome-wide occupancy of a GATA factor in yeast, our rationale was to
120 choose Dal80 as it is known to be highly expressed in derepressing conditions and forms chromosome
121 foci when tagged by GFP (Distler, Kulkarni, Rai, & Cooper, 2001). We thus grew yeast cells in proline-
122 containing media and performed a ChIP-Seq analysis using a Dal80-Myc[13] tagged strain and the
123 isogenic untagged strain, as a control (Fig. 1a). Dal80-bound regions were identified using a peak-
124 calling algorithm (see Material & Methods). A promoter was defined as bound by Dal80 when >75% of
125 the -350 to -100 region (relative to the downstream ORF start site) was overlapped by a peak (Fig. 1b).
126 We chose to use as the reference coordinate the translation initiation codon rather than the
127 transcription start site (TSS) because the latter has not been accurately determined for all genes. Then,
128 our arbitrary definition of the promoter as the -350 to -100 region relative to the ATG codon was based

129 on the distribution of the TSS-ATG distance for genes with an annotated TSS (median and average
130 distance = 58 and 107 bp, respectively; see Supp. Fig. 1a).
131 Strikingly, Dal80 was found to bind to 1269 gene promoters (Fig. 1c-d and Supp. Table 3a). Figure 1E
132 shows a snapshot of ChIP-Seq signals at the promoter of the *FUR4* gene, identified as bound by Dal80.
133 The number of Dal80-bound promoters, corresponding to 22% of all protein-coding genes promoters,
134 is much higher than expected given the hundred target genes that are generally cited for the GATA
135 transcriptional activators (Godard et al., 2007; Scherens et al., 2006), that are presumed to bind to the
136 same sites as Dal80. However, we noted that some peaks (221) overlapped with several promoters
137 (471), mainly of divergent genes (442). In such cases, it is therefore possible that only one of the two
138 divergent promoters is indeed targeted by Dal80, so that the number of Dal80-bound promoters could
139 actually be a little less important than 1269.
140 As expected, genes related to previous nitrogen regulation screens were present within our list (Airoldi
141 et al., 2016; Boer et al., 2007; Boer, De Winde, Pronk, & Piper, 2003; Cox, Pinchak, & Cooper, 1999;
142 Daugherty, Rai, El Berry, & Cooper, 1993; Godard et al., 2007; Scherens et al., 2006; Tai et al., 2005;
143 Wu, Zhang, Hayes, Panoutsopoulou, & Oliver, 2004)(Supp. Table 3b). Among the genes showing Dal80-
144 binding to the promoter, we also noticed a strong enrichment for cytoplasmic translation genes, as
145 well as genes involved in small molecule biosyntheses, including amino acids (Supp. Table 3c).
146 Surprisingly, analysis of GATA site occurrence over Dal80-bound and unbound promoters revealed no
147 major difference between the two classes (Fig. 1f), 48.2% and 51.3% of Dal80-bound and unbound
148 promoters containing at least two GATA sites, respectively (see Supp. Table 3d). Intriguingly, 20% of
149 Dal80-bound promoters do not contain any GATA site (Fig. 1f and Supp. Table 3d), indicating that Dal80
150 recruitment can also occur independently of the presence of consensus GATA sites (see Supp. Fig. 1b
151 for visualization of Dal80 binding to a GATA-less promoter).
152 In summary, our ChIP-Seq analysis revealed that Dal80 binds to a set of promoters larger than
153 previously expected, targeting biosynthetic functions and protein synthesis in addition to nitrogen
154 catabolite repression.

155 **Dal80 binds the intragenic region of a subset of genes**

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

160 We identified 189 genes showing Dal80 intragenic occupancy, according to a >75% overlap of the ORF
161 by a Dal80-Myc[13] peak (Fig. 2a-b). Among them, 144 (76%) were also bound at the promoter (Fig. 2b).
162 Conversely, only 11% of the Dal80-bound promoters (144/1269) showed downstream intragenic
163 binding (Fig. 2b). On the other hand, 45 genes showing Dal80 intragenic binding were not bound at the
164 promoter (Fig. 2b). Hence, we distinguished four classes of genes (Supp. Table 4): (*i*) those bound by
165 Dal80 at the promoter only ("P" class; Fig. 2c), (*ii*) those showing both promoter and intragenic binding
166 ("P & O" class; Fig. 2d), (*iii)* those bound across the ORF only ("O" class; Fig. 2e), (*iv*) the unbound genes
167 (Fig. 2f). Interestingly, we noted that the global Dal80-Myc[13] signal at the promoter was higher for the
168 "P & O" class in comparison to the "P" class (compare Fig. 2c-d).

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

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

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

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

181 Analysis of mRNA levels in proline-grown cells revealed a significant correlation between Dal80 binding
182 and regulation at the mRNA level: 39.6% (216/546) of Dal80-regulated genes were bound by Dal80 at
183 their promoter and/or across the ORF, which is significantly more than expected if Dal80 binding and
184 regulation were independent (P<0.00001, Chi-square test; Supp. Fig. 2a). More importantly, additional
185 differences are emerging between the different classes of genes. First, gene expression was found to
186 be globally similar between the unbound genes and those of the "O" class (P=0.15, Wilcoxon rank-sum
187 test; Fig. 3b). Second, Dal80 binding at the promoter of genes correlated with significantly increased
188 expression levels, in comparison to the unbound genes ($P < 2.2e^{-16}$, Wilcoxon rank-sum test; Fig. 3b).
189 Third, the genes bound by Dal80 both at the promoter and across the ORF ("P & O" class) were even
190 more expressed than the genes showing promoter binding only ($P = 1.3e^{-14}$, Wilcoxon rank-sum test;
191 Fig. 3b). Together with the observation that genes of the "P & O" class globally showed higher Dal80-
192 Myc[13] ChIP-Seq signal at the promoter than those of the "P" class (Fig. 2c-d), our results indicate that
193 Dal80 spreading across gene bodies correlates with a strong recruitment at the promoter, higher
194 sensitivity to Dal80 and high expression in proline-containing medium. Consistently, metagene analysis
195 of Dal80-regulated genes is very similar to that of "P & O" class (Supp. Fig. 2b). Noticeably, 15% (170)
196 and 2% (75) of the genes belonging to the "P" and "unbound" classes respectively display expression
197 levels higher than median levels of expression of the genes belonging to the "P&O" class (Supp. Fig.
198 2c). These proportions raise even further when the first quartile (25%) of the "P&O" class is considered
199 (Supp. Fig. 2c), suggesting that elevated expression levels does not automatically imply binding at ORF.
200 In conclusion, our genome-wide analyses revealed a significant correlation between Dal80 recruitment
201 to the chromatin and Dal80-sensitive expression. We also showed that Dal80 binding both at the
202 promoter and across the gene body correlates with high levels of gene expression.

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

204 In order to validate our genome-wide observations at the level of known NCR-sensitive genes, we
205 characterized the binding profiles of Dal80-Myc[13] along four NCR-sensitive genes of the "P & O" class
206 (see snapshots of ChIP-Seq signals in Fig. 4a-d). We chose the NAD^+-dependent glutamate

207 dehydrogenase gene (*GDH2*) and three permease-coding genes (*GAP1*, *DAL5* and *MEP2*), aware that
208 their regulation could differ (Isabelle Georis, Tate, Cooper, & Dubois, 2011; Godard et al., 2007;
209 Scherens et al., 2006) and that GATA sites are differently represented throughout these genes (Fig. 4a-
210 d).
211 As expected from our genome-wide observations, Dal80 bound not only the promoter, but also the
212 coding region of *DAL5*, *GAP1*, *GDH2* and *MEP2* in proline-grown cells (Fig. 4e and Supp. Fig. 3a). Binding
213 of Dal80 to the *GDH2* ORF was reduced (Fig. 4e), which is consistent with the ChIP-Seq profile (Fig. 4d).
214 We ruled out any erroneous interpretation of our data upon showing that, using the same
215 experimental design, the Gal4-Myc^{13} protein only occupied the *GAL10* promoter, and not the coding
216 region (Supp. Fig. 3b). Moreover, intragenic GATA factor binding did not rely on an intermediate RNA
217 molecule, since the Dal80-Myc^{13} signal across the *MEP2* gene body was unaffected upon RNAse
218 treatment of the chromatin extracts before the immunoprecipitation (Fig. 4f and Supp. Fig. 3c).
219 The results of the ChIP-qPCR show that the level of Dal80 binding also varied from one gene to another.
220 As striking example, Dal80 was more present at the *GAP1* and *MEP2* ORFs, which are more highly
221 expressed than *DAL5* in proline-grown cells (Supp. Table 5a). This is consistent with the correlation
222 observed at the genomic level between Dal80 binding both to the promoter and across gene bodies
223 and high expression levels (Fig. 3b). Altogether, these observations prompted an important
224 mechanistic question: how can Dal80 be localized to gene bodies ?
225 **Dal80 binding within gene bodies occurs independently of intragenic GATA sites**
226 Like most NCR-sensitive genes, the four genes described above are all characterized by stretches of
227 GATA sequences in their promoters (Fig. 4a-d). A possible explanation for Dal80 binding across gene
228 bodies could be that functional, intragenic GATA sequences, directly recruit Dal80. Given the low
229 complexity of the GATA consensus, the probability that a gene is devoid of intragenic GATA site is very
230 low. Indeed, only 10% of all yeast ORFs have no GATA site, and most genes contain an average of 5
231 intragenic GATA sites (Supp. Fig. 3c). However, according to current models, GATA sites should be
232 present in stretches of at least two in order to be functional *in vivo*, which is not the case within *DAL5*,

233 *GDH2*, *GAP1* and *MEP2* ORFs, that contain some GATA sites, but relatively distant from each other.
234 Interestingly, 23 of the 144 P&-O-bound genes did not contain any GATA site in their ORF, hence
235 showing that Dal80 spreading across the body of genes does not necessarly require intragenic GATA
236 sites, at least for these genes.
237 In order to test if the presence of a NCR-sensitive promoter could confer intragenic Dal80 binding
238 across the body of a non-NCR-sensitive gene, devoid of any GATA cluster, we placed the *URA3* ORF
239 under the control of the *MEP2* promoter, at the *MEP2 locus* (P_{MEP2}-*URA3*; Fig. 5a). Only two GATA sites
240 are present in the *URA3* ORF, and separated by 300bp. In this strain, the expression of *URA3* became
241 NCR-sensitive and followed wild type *MEP2* expression (Supp. Fig. 4a), correlating with PolII
242 recruitment over the *URA3* ORF (Supp. Fig. 4b). In these conditions, we observed Dal80-Myc^{13} binding
243 at the promoter of *MEP2* and also across *URA3* (Fig. 5b). Importantly, Dal80 was not detected across
244 *URA3* when it was expressed from its native *locus*, under the control of its promoter (Fig. 5b). These
245 results show that significant recruitment of Dal80 at an ORF region must be preceded by promoter
246 recruitment.
247 Finally, we analyzed Dal80 binding at the *UGA4 locus*, another well-characterized NCR-sensitive gene,
248 which is devoid of intragenic GATA site clusters (Fig. 5c). *UGA4* expression is induced by GABA (γ-
249 aminobutyric acid) and is strongly repressed by Dal80 in the absence of the inducer (André, Hein,
250 Grenson, & Jauniaux, 1993). Therefore, to derepress *UGA4* without inducer and to analyze Dal80
251 binding along *UGA4*, we deleted the repressive C-terminal leucine zipper domain of Dal80 (Dal80ΔLZ-
252 Myc^{13}), known for its homodimerization and heterodimerization with the Gzf3 transcriptional
253 repressor (Coornaert et al., 1992; Svetlov & Cooper, 1998)(Fig. 5d). As a control experiment, *UGA4*
254 expression was quantified in Dal80-Myc^{13} and Dal80ΔLZ-Myc^{13} cells. As expected, *UGA4* mRNA level
255 was very low in Dal80-Myc^{13} cells, whereas deletion of Dal80 leucine zipper led to a strong
256 derepression of *UGA4* (Supp. Fig. 4c). Analysis of PolII occupancy across *UGA4* using ChIP confirmed
257 that elevated mRNA levels correlate with increased transcription (Supp. Fig. 4d). In these conditions,
258 full-length Dal80-Myc^{13} binding was restricted to the *UGA4* promoter, while Dal80ΔLZ-Myc^{13} binding

259 was detected at the promoter and across the body of *UGA4* (Fig. 5e), correlating with the elevated
260 transcript and PolII levels. Surprisingly, the leucine zipper of Dal80 and consequently, its dimerization,
261 needed for *UGA4* repression, were not required for its spreading across *UGA4* gene body. Importantly,
262 these results show that promoter binding is not sufficient to confer intragenic binding, but suggest that
263 transcription activation is required.
264 Together, these data indicate that Dal80 binding at the promoter and across the body of NCR-sensitive
265 genes correlated with expression levels and required transcription activation.
266 **Transcription elongation is necessary for Dal80 binding across *MEP2***
267 We hypothesized that the elongating PolII complex could be responsible for Dal80 spreading beyond
268 Dal80-bound promoters.
269 To assess whether the transcriptionally active PolII complex is necessary for Dal80 binding across the
270 *MEP2* gene, we used an *rpb1-1* strain, known to rapidly involve a cessation of PolII activity when cells
271 are shifted to a restrictive temperature of 37°C (T. S. Kim et al., 2010; Nonet, Scafe, Sexton, & Young,
272 1987). We analyzed Dal80-Myc^{13} binding along *MEP2* in WT and *rpb1-1* cells. When *rpb1-1* cells were
273 shifted at 37°C for 1hr, *MEP2* mRNA and PolII levels showed a 2-fold (Supp. Fig. 5a) and >10-fold
274 decrease (Supp. Fig. 5b), respectively, reflecting the expected elongation defect when *rpb1-1* cells are
275 shifted in non-permissive conditions. In the same conditions, we observed a significant >5-fold
276 reduction of Dal80-Myc^{13} levels across the *MEP2* ORF, while the binding at the promoter was not
277 affected (Fig. 6a-b). This result suggests that Dal80 spreading across the body of NCR-sensitive genes
278 would be a consequence of the PolII-driven elongation process.
279 To get insights into the mechanism by which Dal80 associates to gene bodies, we tested whether it
280 physically interacts with the transcriptionnaly engaged form of PolII (Fig. 6c). Total protein extracts
281 from Dal80-Myc^{13} cells were immunoprecipitated with antibodies directed against the PolII CTD and
282 its phospho-forms Ser2P and Ser5P, respectively characteristic of elongating and initiating PolII forms.
283 All three antibodies enabled effective immunoprecipitation, whereas no antibody and nonspecific
284 antibody controls generated a lower or no signal at all. Thus, Dal80 physically interacts with

285 phosphoforms of the PolIII, suggesting a strong association with PolII engaged in active transcription
286 from initiating to elongating polymerase.
287 Together, our data indicate that Dal80 spreading across the body of NCR-sensitive genes depends on
288 active transcription and that Dal80 interacts with the transcriptioanlly active form of PolII, supporting
289 a model where Dal80 spreading across the body of highly expressed, NCR-sensitive genes might be the
290 result of Dal80-PolII association at post-initiation transcription phases.
291

292 **Discussion**

293 Eukaryotic GATA factors belong to a small but important family of DNA binding proteins involved in
294 development and response to environmental changes in multicellular organisms in unicellular
295 organisms, respectively. In yeast, four GATA factors are involved in Nitrogen Catabolite Repression
296 (NCR), controlling gene expression in response to nitrogen source availability. One of them, the Dal80
297 repressor, itself NCR-sensitive, acts to modulate the intensity of NCR responses.
298 Over the past decade, a number of studies have screened the genome aiming at gathering an inventory
299 of genes regulated by the nitrogen source. Although >500 genes have been shown to be differentially
300 expressed upon change of the nitrogen source (Airoldi et al., 2016; Godard et al., 2007), the list of NCR-
301 sensitive genes was reduced to about 100, based on their sensitivity to GATA factors (Boer et al., 2007,
302 2003; Godard et al., 2007; Scherens et al., 2006), suggesting that the number of Dal80 targets would
303 be situated in that range. Here, our ChIP-Seq analysis identified more than 1200 Dal80-bound
304 promoters, which considerably extends the list of potential Dal80 targets. In fact, the number of Dal80-
305 bound promoters could even have been more important. Indeed, the GATA consensus binding site is
306 rather simple and short, so that in yeast, a total number of 10,000 putative binding sites can be found
307 in all protein-coding gene promoters, 2930 promoters having at least 2 GATA sites, which is thought
308 to be a prerequisite for *in vivo* binding and function of the GATA factors. The difference between the
309 number of promoters with ≥2 GATA sites and the number of Dal80-bound promoters suggests the
310 existence of a selectivity for Dal80 recruitment. This selectivity could rely on promoter architecture

311 and/or chromatin structure, conditioning the requirement for auxiliary DNA binding factors that would
312 stabilize Dal80 at some promoters. Moreover, although we observed a significant correlation between
313 Dal80 binding and regulation, the expression of most of the Dal80-bound genes was not affected in a
314 *dal80Δ* mutant strain. Again, Dal80-dependence for transcribing these genes, as well as their NCR
315 sensitivity, could require the presence of cofactors which are not produced or inactive under the tested
316 growth conditions. In mammals, GATA factors also display an extraordinary complexity in the
317 relationships between binding and expression regulation. Like Dal80, GATA-1 and GATA-2 only occupy
318 a small subset of their abundant binding motif throughout the genome, and the presence of the
319 conserved binding site is insufficient to cause GATA-dependent regulation in most instances
320 (Hasegawa et al., 2016). GATA-1 binding kinetics, stoichiometry and heterogenous complex
321 formations, conditioned by composite promoter architecture, influence its transcriptional activity and
322 hence diversify gene expression profiles (Hasegawa et al., 2016).
323 Given the high conservation at the amino acid level between the DNA binding domain of the four yeast
324 NCR GATA factors, it is likely that they all recognize identical sequences (GATAA, GATAAG or GATTAG).
325 This consensus has been largely validated in the past using gene reporter experiments, mutational
326 analyses and *in vitro* binding experiments on naked DNA. Nonetheless, of the 1269 bound promoters,
327 48% contained at least two GATA sites, a proportion that is not different from that observed among
328 unbound promoters, and the amount of GATA sites per promoter was not different between the two
329 groups either. In addition, Dal80 recruitment was found to occur independently of the presence of
330 GATA sites in 20% of Dal80-bound promoters, as also previously observed in mammalian cells (Cheng
331 et al., 2009; Fujiwara et al., 2009). Future experiments will be required to decipher the molecular
332 mechanism by which Dal80 can be recruited to these GATA-less promoters.
333 Unexpectedly, although Dal80 has always been described as a repressor, we identified 314 genes that
334 are positively regulated by Dal80 (their expression is significantly decreased upon Dal80 deletion).
335 These genes are distinct from the NCR-sensitive genes : only 4 belong to the NCR-sensitive list of
336 (Godard et al., 2007) and are significantly enriched in amino acid biosynthetic processes, resembling

337 the amino acid starvation response mediated by the Gcn4 transcriptional activator. Interestingly, the
338 promoter of 122/314 Dal80-activated genes contain Gcn4-binding sites (Supp. Fig. 3c), and this group
339 of 314 Dal80-activated genes is significantly enriched for genes regulated by the General Amino Acid
340 Control (GAAC; YeastMine Gene List, Publication Enrichment, p<1.6e-13), through the Gcn4 activator.
341 Interconnections between NCR and GAAC have already been demonstrated, mostly at the level of
342 nitrogen catabolism control: 1-a large number of non-preferential nitrogen sources leads to increased
343 transcription of GAAC targets (Godard et al., 2007); and 2- Gcn4 contributes, with Gln3, to the
344 expression of NCR-sensitive genes (Staschke et al., 2010; Valenzuela, Aranda, & González, 2001).
345 However, this is the first time that evidence are provided suggesting a positive function for the negative
346 GATA factor at the level biosynthetic gene expression.
347 The most striking and unexpected finding of this work is the observation that for 189 genes, Dal80
348 occupied the gene body ("O" and "P & O" classes). The 45 genes displaying Dal80 intragenic occupancy
349 with no recruitment at the promoter ("O" class) were globally neither Dal80-sensitive nor NCR-
350 sensitive, similar to the unbound genes. Furthermore, we also noted that a substantial fraction of these
351 genes correspond to small dubious ORFs, near or even overlapping the promoter, Dal80-bound, of the
352 adjacent gene. In these cases, the limited resolution of the ChIP-Seq technique, combined to the small
353 size of these genes, might have allowed them to pass through the filters of the procedure to identify
354 the genes showing Dal80 intragenic binding. Overall, these observations suggest that the existence of
355 the "O" class is likely to be physiologically irrelevant. Dal80 binding at the promoter and spreading
356 sacross the body of the 144 genes belonging to the "P & O" class correlated with high expression levels
357 and sensitivity to Dal80-dependent repression. The observation that a considerable fraction of "P"
358 class genes have similar or higher expression levels than most of the "P&O" class genes suggest that
359 the well-known artifact of hyper-chippable regions is not applicable here (Teytelman, Thurtle, Rine, &
360 van Oudenaarden, 2013). Our observations were also experimentally confirmed using ChIP
361 experiments, at the level of well-characterized NCR-sensitive genes. Our data indicate that Dal80
362 spreading across genes bodies does not rely on intragenic GATA sites. Rather, promoter binding

363 appears to be required but not sufficient: Dal80 spreading across genes depends on active
364 transcription. Indeed, the observation that impaired PolII elongation correlates with decreased
365 intragenic binding further indicates that Dal80 binding across gene bodies depends on ongoing
366 transcription elongation. Consistently, we detected a physical interaction between Dal80 and
367 transcriptionnaly active forms of PolII. Together, our data leads us to propose a model where Dal80
368 could travel from the promoter of highly expressed, NCR-sensitive genes through the gene body by
369 accompanying the elongating PolII complex (Figure 7). Additional investigations will be required to
370 define which domain of Dal80 is responsible for the interaction with the transcription machinery, to
371 determine whether there is any causal relationship between Dal80 intragenic binding and high
372 expression levels, and to decipher the potential role of Dal80 during transcription elongation.
373 Whereas the binding of elongation factors across gene bodies has been thoroughly documented (Sims,
374 Belotserkovskaya, & Reinberg, 2004), it has also been described for few specific transcription factors.
375 For example, Gal4 was reported to bind to its consensus DNA target within the *ACC1* ORF, but the
376 authors concluded that the observed transcriptional repression of the *ACC1* gene was most likely
377 resulting from random *GAL4* binding "noise" over the genome, thus having no physiological
378 explanation for this ORF-bound transcription factor (Li & Johnston, 2001). Likewise, Gcn4 was detected
379 across the *PHO8* ORF, with concomitant recruitment of the SAGA complex, but without any impact on
380 gene expression (Topalidou & Thireos, 2003). More recently, binding of the Gcn4 transcription factor
381 to its consensus site at some ORFs, when located in proximity of the transcriptional start site, was
382 found to play a consistent role in controlling embedded cryptic promoters in yeast, thereby affecting
383 Gcn4-dependent transcription of some genes (Rawal et al., 2018).
384 A recent study has identified CTD phosphorylation of polII as a hub that optimizes transcriptome
385 changes to adequately balance optimal growth and stress tolerance responses (Chasman et al., 2014).
386 The addition of nitrogen to nitrogen-limited cells rapidly results in the transient overproduction of
387 transcripts required for protein translation (stimulated growth) whereas accelerated mRNA
388 degradation favours rapid clearing of the most abundant transcripts, like those involved in high affinity

389 permease production, that are highly expressed NCR-sensitive genes, for example (Airoldi et al., 2016).
390 The involvement of the Nrd1-Nab3-Sen1 (NNS) and TRAMP complexes in these regulatory responses
391 has been envisioned very recently (Bresson, Tuck, Staneva, & Tollervey, 2017; Merran & Corden, 2017).
392 Thus, a possible role of Dal80 (and possibly of the other GATA factors) binding along highly expressed
393 genes could be to transmit nutritional signals to the RNA decay machinery for rapid and efficient
394 remodeling of gene expression in response to changing nitrogen supplies. Here, an attractive model
395 (Figure 7) would hypothesize that Dal80 binding to polII could prevent the decay of the dedicated
396 stress-response mRNAs to enhance their cellular life-cycle by regulating/repressing RNA decay factors
397 that might associate otherwise on the polymerase. Other elongation-related processes, like histone
398 modification and chromatin remodeling (Henikoff & Shilatifard, 2011; Tessarz & Kouzarides, 2014),
399 mRNA export (Meinel et al., 2013) or roadblock termination (Candelli et al., 2018) can also be
400 proposed.
401 In this respect, in human cells, GATA factors are also reported to occupy non-canonical sites within the
402 genome, further reinforcing that they can be recruited to the chromatin independently of their motif
403 (Cheng et al., 2009; Fujiwara et al., 2009). In addition, 43% of the GATA1 peaks were collected among
404 exon, introns and 3'UTR of coding genes in human erythroleukemia cells (Fujiwara et al., 2009). It is
405 tempting to hypothesize that GATA factors could have a dual or synergistic role during transcription,
406 i.e. recruiting/stabilizing the PIC complex as for any classical transcription factor in the
407 promoter/enhancer regions and promoting elongation at a post initiation step interacting with the
408 RNAPII. Interestingly, GATA1 in erythropoid cells interacts with the mediator (Stumpf et al., 2006), a
409 complex also known to travel with PolII (Allen & Taatjes, 2015).
410

411 ## Materials and Methods

412 **Experimental Model and Subject Details**

413 Experiments were conducted using *S. cerevisiae* strains of the FY genetic background. The strains used
414 are listed in Supp. Table 1. Dal80 was tagged with 13 copies of the *c-myc* epitope (Myc^{13}) as described

415 (Longtine et al., 1998) using primers listed in Supp. Tables 1 and 2. The P_{MEP2}-*URA3* allele in strains
416 FV806-808 was created by amplification of the *URA3* gene using the same strategy, with primers listed
417 in Supp. Tables 1 and 2.
418 Cultures were grown at 29°C to mid-log phase (A_{660nm} = 0.5) in YNB (without amino acids or ammonia)
419 minimal medium containing the indicated nitrogen source at a 0.1% final concentration, glucose (3%)
420 and the appropriate supplements (20 µg/ml uracil, histidine and tryptophan) to cover auxotrophic
421 requirements.
422 **Chromatin immunoprecipitation**
423 Cell extracts and immunoprecipitations were conducted as described (I. Georis et al., 2009) using
424 primers listed in Supp. Table 2. The cells (100-ml cultures grown to an absorbance (A660 nm = 0.6)
425 corresponding to 6 × 106 cells/ml) were treated with 1% formaldehyde for 30 min at 25 °C and mixed
426 by orbital shaking. Glycine was then added to a final concentration of 500 mm and incubation
427 continued for 5 min. The cells were collected, washed once with cold 10 mm Tris-HCl, pH 8, washed
428 once with cold FA-SDS buffer (50 mm HEPES-KOH, pH 7.5, 150 mm NaCl, 1 mm EDTA, 1% Triton X-100,
429 0.1% sodium deoxycholate, 0.1% SDS, 1 mm phenylmethylsulfonyl fluoride), and resuspended in 1 ml
430 of cold FA-SDS buffer. An equal volume of glass beads (0.5 mm in diameter) was added, and the cells
431 were disrupted by vortexing for 30 min in a cold room. The lysate was diluted into 4 ml of FA-SDS
432 buffer, and the glass beads were discarded. The cross-linked chromatin was then pelleted by
433 centrifugation (17,000 × g for 35 min), washed for 60 min with FA-SDS buffer, resuspended in 1.6 ml
434 of FA-SDS buffer for 15 min at 4 °C, and sonicated three times for 30 s. each (Bioruptor, Diagenode) to
435 yield an average DNA fragment size of 700 base pairs. Finally, the sample was clarified by centrifugation
436 at 14,000 × g for 30 min and diluted 4-fold in FA-SDS buffer, and aliquots of the resultant chromatin
437 containing solution were stored at –80 °C. Myc^{13}-tagged proteins were immunoprecipitated by
438 incubating 100 µl of the chromatin containing solution for 180 min at 4°C with 2 µl of mouse anti-Myc
439 antibodies (SC-40, Santa Cruz) prebound to 10 µl of Dynabeads Pan Mouse IgG (Dynal) according to
440 the manufacturer's instructions. Immune complexes were washed six times in FA-SDS buffer and

441 recovered by treating with 50 µl of Pronase Buffer (25 mm Tris, pH 7.5, 5 mm EDTA, 0.5% SDS) at 65
442 °C with agitation. Input (IN) and immunoprecipitated (IP) fractions were then subjected to Pronase
443 treatment (0.5 mg/ml; Roche Applied Science) for 60 min at 37°C, and formaldehyde cross-links were
444 reversed by incubating the eluates overnight at 65°C. Finally, the samples were treated with RNase (50
445 µg/ml) for 60 min at 37°C. DNA from the IP fractions was purified using the High Pure PCR Product
446 Purification Kit (Roche Applied Science) and eluted in 50 µl of 20 mm Tris buffer, pH 8. IN fractions
447 were boiled 10 min and diluted 500-fold with no further purification prior to quantitative PCR analysis.

448 **Quantitative RT-PCR**

449 Quantitative RT-PCR was performed as described previously (I. Georis et al., 2009) using primers listed
450 in Supp. Table 2. Total RNA was extracted from 4-ml cultures and cDNA was generated from 100 to
451 500 ng of total RNA using a RevertAid H Minus first-strand cDNA synthesis kit with oligo$(dT)_{18}$ primers
452 from Fermentas using the manufacturer's recommended protocol. cDNAs were subsequently
453 quantified by RT-PCR using the Maxima SYBR green qPCR master mix from Fermentas.

454 **Co-Immunoprecipitation**

455 Cultures (100 ml) were harvested, washed once in 50 mM Tris, pH 8, and resuspended in 1ml of buffer
456 (50 mM Tris, pH 8, 150 mM NaCl, 5 mM EDTA, 0.05% NP-40, 1 mM phenylmetyhlsulfonyl fluoride, and
457 complete protease inhibitor cocktail tablets [Roche]). Lysis was performed by shaking with 425-600
458 µm acid-washed glass beads (Sigma) on an IKA Vibrax VXR orbital shaker at maximum speed for 30 min
459 at 4°C. Cell debris and glass beads were removed by centrifugation. Immunoprecipitation was
460 performed by incubating 200 µl of total cell extracts with 20 µl of Dynabeads PAN mouse
461 immunoglobulin G (Invitrogen) that were preinucbated with anti-HA (SCBT, SC-7392), anti-CTD (SCBT,
462 CTD4H8), anti-Ser2P (BioLegend, H5) or anti-Ser5P (BioLegend, H14) antibodies and 20 µl of 1%
463 phosphate-buffered saline-bovine serum albumin for 2 h under orbital shaking (800 rpm) at 30°C.
464 Immune complexes were washed three times in lysis buffer, eluted by boiling in sodium dodecyl sulfate
465 (SDS) sample buffer, and loaded on SDS-polyacrylamide gel for anti-Myc Western blotting.

466 **ChIP-Seq analysis and Peak-calling**

467 ChIP-Seq analysis was performed from two biological replicates of proline-grown 25T0b (no tag) and
468 FV078 (Dal80-Myc[13]) cells. For each condition, libraries were prepared from 10 ng of "input" or "IP"
469 DNA using the TruSeq ChIP Sample Preparation Kit (Illumina). Single-read sequencing (50 nt) of the
470 libraries was performed on a HiSeq 2500 sequencer.
471 Reads were uniquely mapped to the *S. cerevisiae* S288C reference genome using Bowtie2 v2.1.0
472 (Langmead & Salzberg, 2012), with a tolerance of 1 mismatch in seed alignment. Tags densities were
473 normalized on the total number of uniquely reads mapped.
474 Dal80-bound regions were identified through a peak-calling procedure using version 2.0.9 of MACS
475 (Zhang et al., 2008), with a minimum false discovery rate (FDR) of 0.001.
476 **Total RNA-Seq**
477 For each strain and condition, total RNA was extracted from two biological replicates using standard
478 hot phenol procedure, ethanol-precipitated, resuspended in nuclease-free H_2O (Ambion) and
479 quantified using a NanoDrop 2000c spectrophotometer. Ribosomal RNAs were depleted from 1 µg of
480 total RNA using the RiboMinus™ Eukaryote v2 Kit (Life Technologies). After concentration using the
481 Ribominus™ Concentration Module (Life Technologies), rRNA-depleted RNA was quantified using the
482 Qubit RNA HS Assay kit (Life Technologies). In parallel, rRNA depletion efficiency and integrity of both
483 total and rRNA-depleted RNA were checked by analysis in a RNA 6000 Pico chip, in a 2100 bioanalyzer
484 (Agilent). Strand-specific total RNA-Seq libraries were prepared from 125 ng of rRNA-depleted RNA
485 using the TruSeq® Stranded Total RNA Sample Preparation Kit (Illumina), following manufacturer's
486 instructions. Paired-end sequencing (2 x 50 nt) of the libraries was performed on a HiSeq 2500
487 sequencer. Sequenced reads were mapped to the reference genome using version 2.0.6 of TopHat (D.
488 Kim et al., 2013), as described (Wery et al., 2016). Tags densities were normalized on the total number
489 of reads uniquely mapped on ORFs. Differential expression analysis was performed using DESeq
490 (Anders & Huber, 2010).

491 **Quantification and Statistical Analysis**

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

495 **Availability of data and materials**

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

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

502

503 ## Acknowledgements

504 This work has benefited from the facilities and expertise of the NGS platform of Institut Curie,
505 supported by the Agence Nationale de la Recherche (ANR-10-EQPX-03, ANR10-INBS-09-08) and the
506 Canceropôle Ile-de-France. A. Morillon's lab is supported by the Agence Nationale de la Recherche
507 (DNA-Life) and the European Research Council (EpincRNA starting grant, DARK consolidator grant).
508 Work by IG and ED is funded by the CoCoF and FNRS (FRFC 2.4547.11). Work by AR was funded by a
509 FRIA fellowship.

510 We are grateful to all members of our labs for discussions and critical reading of the manuscript. We
511 thank F. Vierendeels for excellent technical assistance.

512

513 ## Competing financial interest

514 The authors declare that they have no competing interests.

515

516 ## References

517 Airoldi, E. M., Miller, D., Athanasiadou, R., Brandt, N., Abdul-Rahman, F., Neymotin, B., … Gresham,
518 D. (2016). Steady-state and dynamic gene expression programs in Saccharomyces cerevisiae in
519 response to variation in environmental nitrogen. *Molecular Biology of the Cell*, *27*, 1–38.
520 https://doi.org/10.1091/mbc.E14-05-1013

521 Allen, B. L., & Taatjes, D. J. (2015). The Mediator complex: a central integrator of transcription.
522 *Nature Reviews Molecular Cell Biology*, *16*(3), 155–166. https://doi.org/10.1038/nrm3951

523 Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome*
524 *Biology*, *11*(10), R106. https://doi.org/10.1186/gb-2010-11-10-r106

525 André, B., Hein, C., Grenson, M., & Jauniaux, J. C. (1993). Cloning and expression of the UGA4 gene
526 coding for the inducible GABA-specific transport protein of Saccharomyces cerevisiae.
527 *Molecular & General Genetics*, *237*(1–2), 17–25. https://doi.org/10.1007/BF00282779

528 André, B., Talibi, D., Boudekou, S. S., Hein, C., Vissers, S., & Coornaert, D. (1995). Two mutually
529 exclusive regulatory systems inhibit UASGATA, a cluster of 5'-GAT(A/T)A-3' upstream from the
530 UGA4 gene of Saccharomyces cerevisiae. *Nucleic Acids Research*, *23*(4), 558–564.
531 https://doi.org/10.1093/nar/23.4.558

532 Boer, V. M., De Winde, J. H., Pronk, J. T., & Piper, M. D. W. (2003). The genome-wide transcriptional
533 responses of Saccharomyces cerevisiae grown on glucose in aerobic chemostat cultures limited
534 for carbon, nitrogen, phosphorus, or sulfur. *Journal of Biological Chemistry*, *278*(5), 3265–3274.
535 https://doi.org/10.1074/jbc.M209759200

536 Boer, V. M., Tai, S. L., Vuralhan, Z., Arifin, Y., Walsh, M. C., Piper, M. D. W., … Daran, J. M. (2007).
537 Transcriptional responses of Saccharomyces cerevisiae to preferred and nonpreferred nitrogen
538 sources in glucose-limited chemostat cultures. *FEMS Yeast Research*, *7*(4), 604–620.
539 https://doi.org/10.1111/j.1567-1364.2007.00220.x

540 Bresnick, E. H., Katsumura, K. R., Lee, H. Y., Johnson, K. D., & Perkins, A. S. (2012). Master regulatory
541 GATA transcription factors: Mechanistic principles and emerging links to hematologic

542 malignancies. *Nucleic Acids Research*, *40*(13), 5819–5831. https://doi.org/10.1093/nar/gks281
543 Bresnick, E. H., Lee, H. Y., Fujiwara, T., Johnson, K. D., & Keles, S. (2010). GATA switches as
544 developmental drivers. *Journal of Biological Chemistry*, *285*(41), 31087–31093.
545 https://doi.org/10.1074/jbc.R110.159079
546 Bresson, S., Tuck, A., Staneva, D., & Tollervey, D. (2017). Nuclear RNA Decay Pathways Aid Rapid
547 Remodeling of Gene Expression in Yeast. *Molecular Cell*, *65*(5), 787–800.
548 https://doi.org/10.1016/j.molcel.2017.01.005
549 Bysani, N., Daugherty, J. R., & Cooper, T. G. (1991). Saturation mutagenesis of the UAS(NTR) (GATAA)
550 responsible for nitrogen catabolite repression-sensitive transcriptional activation of the
551 allantoin pathway genes in Saccharomyces cerevisiae. *Journal of Bacteriology*, *173*(16), 4977–
552 4982.
553 Candelli, T., Challal, D., Briand, J., Boulay, J., Porrua, O., Colin, J., & Libri, D. (2018). High-resolution
554 transcription maps reveal the widespread impact of roadblock termination in yeast. *The EMBO*
555 *Journal*, *37*(4), e97490. https://doi.org/10.15252/embj.201797490
556 Chasman, D., Ho, Y.-H., Berry, D. B., Nemec, C. M., MacGilvray, M. E., Hose, J., … Gasch, A. P. (2014).
557 Pathway connectivity and signaling coordination in the yeast stress-activated signaling network.
558 *Molecular Systems Biology*, *10*(11), 759. https://doi.org/10.15252/msb.20145120
559 Cheng, Y., Wu, W., Kumar, S. A., Yu, D., Deng, W., Tripic, T., … Hardison, R. C. (2009). Erythroid GATA1
560 function revealed by genome-wide analysis of transcription factor occupancy, histone
561 modifications, and mRNA expression. *Genome Research*, *19*(12), 2172–2184.
562 https://doi.org/10.1101/gr.098921.109
563 Coffman, J. A., Rai, R., Cunningham, T., Svetlov, V., & Cooper, T. G. (1996). Gat1p, a GATA family
564 protein whose production is sensitive to nitrogen catabolite repression, participates in
565 transcriptional activation of nitrogen-catabolic genes in Saccharomyces cerevisiae. *Molecular*
566 *and Cellular Biology*, *16*(3), 847–858.
567 Coffman, J. A., Rai, R., Loprete, D. M., Cunningham, T., Svetlov, V., & Cooper, T. G. (1997). Cross

568 regulation of four GATA factors that control nitrogen catabolic gene expression in
569 Saccharomyces cerevisiae. *Journal of Bacteriology*, *179*(11), 3416–3429.
570 Cooper, T. G., Rai, R., & Yoo, H. S. (1989). Requirement of upstream activation sequences for nitrogen
571 catabolite repression of the allantoin system genes in Saccharomyces cerevisiae. *Molecular and*
572 *Cellular Biology*, *9*(12), 5440–5444.
573 Coornaert, D., Vissers, S., Andre, B., & Grenson, M. (1992). The UGA43 negative regulatory gene of
574 Saccharomyces cerevisiae contains both a GATA-1 type zinc finger and a putative leucine zipper.
575 *Current Genetics*, *21*(0172–8083 (Print)), 301–307.
576 Cox, K. H., Pinchak, A. B., & Cooper, T. G. (1999). Genome-wide transcriptional analysis in S.
577 cerevisiae by mini-array membrane hybridization. *Yeast*, *15*(8), 703–713.
578 https://doi.org/10.1002/(SICI)1097-0061(19990615)15:8<703::AID-YEA413>3.0.CO;2-Z
579 Cunningham, T. S., & Cooper, T. G. (1991). Expression of the DAL80 gene, whose product is
580 homologous to the GATA factors and is a negative regulator of multiple nitrogen catabolic
581 genes in Saccharomyces cerevisiae, is sensitive to nitrogen catabolite repression. *Molecular and*
582 *Cellular Biology*, *11*(12), 6205–15. https://doi.org/10.1128/MCB.11.12.6205
583 Cunningham, T. S., & Cooper, T. G. (1993). The Saccharomyces cerevisiae DAL80 repressor protein
584 binds to multiple copies of GATAA-containing sequences (URS(GATA)). *Journal of Bacteriology*,
585 *175*(18), 5851–5861. https://doi.org/10.1128/JB.175.18.5851-5861.1993
586 Cunningham, T. S., Dorrington, R. A., & Cooper, T. G. (1994). The UGA4 UAS(NTR) site required for
587 GLN3-dependent transcriptional activation also mediates DAL80-responsive regulation and
588 DAL80 protein binding in Saccharomyces cerevisiae. *Journal of Bacteriology*, *176*(15), 4718–
589 4725.
590 Cunningham, T. S., Svetlov, V. V., Rai, R., Smart, W., & Cooper, T. G. (1996). Gln3p Is capable of
591 binding to UASNTR elements and activating transcription in Saccharomyces cerevisiae. *Journal*
592 *of Bacteriology*, *178*(12), 3470–3479.
593 Daugherty, J. R., Rai, R., El Berry, H. M., & Cooper, T. G. (1993). Regulatory circuit for responses of

594 nitrogen catabolic gene expression to the GLN3 and DAL80 proteins and nitrogen catabolite
595 repression in Saccharomyces cerevisiae. *Journal of Bacteriology*, *175*(1), 64–73.
596 Descrimes, M., Zouari, Y. Ben, Wery, M., Legendre, R., Gautheret, D., & Morillon, A. (2015). VING: a
597 software for visualization of deep sequencing signals. *BMC Research Notes*, *8*(1), 419.
598 https://doi.org/10.1186/s13104-015-1404-5
599 Distler, M., Kulkarni, A., Rai, R., & Cooper, T. G. (2001). Green fluorescent protein-Dal80p illuminates
600 up to 16 distinct foci that colocalize with and exhibit the same behavior as chromosomal DNA
601 proceeding through the cell cycle of saccharomyces cerevisiae. *Journal of Bacteriology*, *183*(15),
602 4636–4642. https://doi.org/10.1128/JB.183.15.4636-4642.2001
603 Fujiwara, T., O'Geen, H., Keles, S., Blahnik, K., Linnemann, A. K., Kang, Y. A., … Bresnick, E. H. (2009).
604 Discovering Hematopoietic Mechanisms through Genome-wide Analysis of GATA Factor
605 Chromatin Occupancy. *Molecular Cell*, *36*(4), 667–681.
606 https://doi.org/10.1016/j.molcel.2009.11.001
607 Georis, I., Feller, A., Vierendeels, F., & Dubois, E. (2009). The Yeast GATA Factor Gat1 Occupies a
608 Central Position in Nitrogen Catabolite Repression-Sensitive Gene Activation. *Molecular and*
609 *Cellular Biology*, *29*(13), 3803–3815. https://doi.org/10.1128/MCB.00399-09
610 Georis, I., Tate, J. J., Cooper, T. G., & Dubois, E. (2011). Nitrogen-responsive regulation of GATA
611 protein family activators Gln3 and Gat1 occurs by two distinct pathways, one inhibited by
612 rapamycin and the other by methionine sulfoximine. *Journal of Biological Chemistry*, *286*(52),
613 44897–44912. https://doi.org/10.1074/jbc.M111.290577
614 Georis, I., Tate, J. J., Vierendeels, F., Cooper, T. G., & Dubois, E. (2015). Premature termination of
615 GAT1 transcription explains paradoxical negative correlation between nitrogen-responsive
616 mRNA, but constitutive low-level protein production. *RNA Biology*, *12*(8), 824–837.
617 https://doi.org/10.1080/15476286.2015.1058476
618 Godard, P., Urrestarazu, A., Vissers, S., Kontos, K., Bontempi, G., van Helden, J., & Andre, B. (2007).
619 Effect of 21 Different Nitrogen Sources on Global Gene Expression in the Yeast Saccharomyces

620 cerevisiae. *Molecular and Cellular Biology*, *27*(8), 3065–3086.
621 https://doi.org/10.1128/MCB.01084-06
622 Grass, J. A., Boyer, M. E., Pal, S., Wu, J., Weiss, M. J., & Bresnick, E. H. (2003). GATA-1-dependent
623 transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide
624 chromatin remodeling. *Proceedings of the National Academy of Sciences of the United States of*
625 *America*, *100*(15), 8811–8816. https://doi.org/10.1073/pnas.1432147100
626 Hahn, S., & Young, E. T. (2011). Transcriptional regulation in Saccharomyces cerevisiae: Transcription
627 factor regulation and function, mechanisms of initiation, and roles of activators and
628 coactivators. *Genetics*, *189*(3), 705–736. https://doi.org/10.1534/genetics.111.127019
629 Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., … Young, R. A.
630 (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature*, *431*(7004), 99–104.
631 https://doi.org/10.1038/nature02800
632 Hasegawa, A., Kaneko, H., Ishihara, D., Nakamura, M., Watanabe, A., Yamamoto, M., … Shimizu, R.
633 (2016). GATA1 binding kinetics on conformation-specific binding sites elicit differential
634 transcriptional regulation. *Molecular and Cellular Biology*, *36*(16), 2151–2167.
635 https://doi.org/10.1128/MCB.00017-16
636 Henikoff, S., & Shilatifard, A. (2011). Histone modification: Cause or cog? *Trends in Genetics*, *27*(10),
637 389–396. https://doi.org/10.1016/j.tig.2011.06.006
638 Horak, C. E., Mahajan, M. C., Luscombe, N. M., Gerstein, M., Weissman, S. M., & Snyder, M. (2002).
639 GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis.
640 *Proceedings of the National Academy of Sciences of the United States of America*, *99*(5), 2924–
641 2929. https://doi.org/10.1073/pnas.052706999
642 Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate
643 alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome*
644 *Biology*, *14*(4), R36. https://doi.org/10.1186/gb-2013-14-4-r36
645 Kim, T. S., Liu, C. L., Yassour, M., Holik, J., Friedman, N., Buratowski, S., & Rando, O. J. (2010). RNA

646 polymerase mapping during stress responses reveals widespread nonproductive transcription in
647 yeast. *Genome Biology*, *11*(7), R75. https://doi.org/10.1186/gb-2010-11-7-r75
648 Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*,
649 *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923
650 Lee, T. I. (2002). Transcriptional Regulatory Networks in Saccharomyces cerevisiae. *Science*,
651 *298*(5594), 799–804. https://doi.org/10.1126/science.1075090
652 Levine, M., & Tjian, R. (2003). Transcription regulation and animal diversity. *Nature*, *424*(6945), 147–
653 151. https://doi.org/10.1038/nature01763
654 Li, Q., & Johnston, S. A. (2001). Are All DNA Binding and Transcription Regulation by an Activator
655 Physiologically Relevant? *Molecular and Cellular Biology*, *21*(7), 2467–2474.
656 https://doi.org/10.1128/MCB.21.7.2467-2474.2001
657 Longtine, M. S., McKenzie, A., Demarini, D. J., Shah, N. G., Wach, A., Brachat, A., … Pringle, J. R.
658 (1998). Additional modules for versatile and economical PCR-based gene deletion and
659 modification in Saccharomyces cerevisiae. *Yeast*, *14*(10), 953–961.
660 https://doi.org/10.1002/(SICI)1097-0061(199807)14:10<953::AID-YEA293>3.0.CO;2-U
661 Lowry, J. A., & Atchley, W. R. (2000). Molecular Evolution of the GATA Family of Transcription
662 Factors: Conservation Within the DNA-Binding Domain. *Journal of Molecular Evolution*, *50*(2),
663 103–115. https://doi.org/10.1007/s002399910012
664 MacIsaac, K. D., Wang, T., Gordon, D. B., Gifford, D. K., Stormo, G. D., & Fraenkel, E. (2006). An
665 improved map of conserved regulatory sites for Saccharomyces cerevisiae. *BMC Bioinformatics*,
666 *7*, 113. https://doi.org/10.1186/1471-2105-7-113
667 Magasanik, B., & Kaiser, C. A. (2002). Nitrogen regulation in Saccharomyces cerevisiae. *Gene*, *290*(1–
668 2), 1–18. https://doi.org/10.1016/S0378-1119(02)00558-9
669 Meinel, D. M., Burkert-Kautzsch, C., Kieser, A., O'Duibhir, E., Siebert, M., Mayer, A., … Sträßer, K.
670 (2013). Recruitment of TREX to the Transcription Machinery by Its Direct Binding to the
671 Phospho-CTD of RNA Polymerase II. *PLoS Genetics*, *9*(11), e1003914.

https://doi.org/10.1371/journal.pgen.1003914

Merran, J., & Corden, J. L. (2017). Yeast RNA-Binding Protein Nab3 Regulates Genes Involved in Nitrogen Metabolism. *Mol. Cell. Biol.*, *37*(18), 1–11.

Minehart, P. L., & 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*(12), 6216–6228. https://doi.org/10.1128/MCB.11.12.6216

Nonet, M., Scafe, C., Sexton, J., & Young, R. (1987). Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. *Molecular and Cellular Biology*, *7*(5), 1602–1611. https://doi.org/10.1128/MCB.7.5.1602.Updated

Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., … 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*(5120), 438–446. https://doi.org/10.1126/science.8332909

Rai, R., Genbauffe, F. S., Sumrada, R. A., & Cooper, T. G. (1989). Identification of sequences responsible for transcriptional activation of the allantoate permease gene in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *9*(2), 602–608.

Rawal, Y., Chereji, R., Valabhoju, V., Qiu, H., Ocampo, J., Clark, D., & Hinnebusch, A. G. (2018). Gcn4 Binding in Coding Regions Can Activate Internal and Canonical 5' Promoters in Yeast. *Molecular Cell*, *70*(2), 297–311. https://doi.org/10.1016/j.molcel.2018.03.007

Rowen, D. W., Esiobu, N., & Magasanik, B. (1997). Role of GATA factor Nil2p in nitrogen regulation of gene expression in Saccharomyces cerevisiae. *Journal of Bacteriology*, *179*(11), 3761–3766.

Scherens, B., Feller, A., Vierendeels, F., Messenguy, F., & Dubois, E. (2006). Identification of direct and indirect targets of the Gln3 and Gat1 activators by transcriptional profiling in response to nitrogen availability in the short and long term. *FEMS Yeast Research*, *6*(5), 777–791. https://doi.org/10.1111/j.1567-1364.2006.00060.x

698 Sims, R. J., Belotserkovskaya, R., & Reinberg, D. (2004). Elongation by RNA polymerase II: The short
699 and long of it. *Genes and Development*, *18*(20), 2437–2468.
700 https://doi.org/10.1101/gad.1235904
701 Soussi-Boudekou, S., Vissers, S., Urrestarazu, A., Jauniaux, J. C., & André, B. (1997). Gzf3P, a fourth
702 GATA factor involved in nitrogen-regulated transcription in Saccharomyces cerevisiae.
703 *Molecular Microbiology*, *23*(6), 1157–1168. https://doi.org/10.1046/j.1365-
704 2958.1997.3021665.x
705 Stanbrough, M., & Magasanik, B. (1996). Two transcription factors, Gln3p and Nil1p, use the same
706 GATAAG sites to activate the expression of GAP1 of Saccharomyces cerevisiae. *Journal of*
707 *Bacteriology*, *178*(8), 2465–2468.
708 Starich, M. R., Wikström, M., Arst, H. N., Clore, G. M., & Gronenborn, A. M. (1998). The solution
709 structure of a fungal AREA protein-DNA complex: an alternative binding mode for the basic
710 carboxyl tail of GATA factors. *Journal of Molecular Biology*, *277*(3), 605–620.
711 https://doi.org/10.1006/jmbi.1998.1625
712 Staschke, K. A., Dey, S., Zaborske, J. M., Palam, L. R., McClintick, J. N., Pan, T., … Wek, R. C. (2010).
713 Integration of general amino acid control and Target of Rapamycin (TOR) regulatory pathways
714 in nitrogen assimilation in yeast. *Journal of Biological Chemistry*, *285*(22), 16893–16911.
715 https://doi.org/10.1074/jbc.M110.121947
716 Stumpf, M., Waskow, C., Krötschel, M., van Essen, D., Rodriguez, P., Zhang, X., … Borggrefe, T. (2006).
717 The mediator complex functions as a coactivator for GATA-1 in erythropoiesis via subunit
718 Med1/TRAP220. *Proceedings of the National Academy of Sciences of the United States of*
719 *America*, *103*(49), 18504–18509. https://doi.org/10.1073/pnas.0604494103
720 Svetlov, V. V., & Cooper, T. G. (1998). The Saccharomyces cerevisiae GATA factors Dal80p and Deh1p
721 can form homo- and heterodimeric complexes. *Journal of Bacteriology*, *180*(21), 5682–5688.
722 Tai, S. L., Boer, V. M., Daran-Lapujade, P., Walsh, M. C., de Winde, J. H., Daran, J. M., & Pronk, J. T.
723 (2005). Two-dimensional transcriptome analysis in chemostat cultures. *J Biol Chem*, *280*(1),

437–447.

Tessarz, P., & Kouzarides, T. (2014). Histone core modifications regulating nucleosome structure and dynamics. *Nature Reviews Molecular Cell Biology*, *15*(11), 703–708. https://doi.org/10.1038/nrm3890

Teytelman, L., Thurtle, D. M., Rine, J., & van Oudenaarden, A. (2013). Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proceedings of the National Academy of Sciences*, *110*(46), 18602–18607. https://doi.org/10.1073/pnas.1316064110

Topalidou, I., & Thireos, G. (2003). Gcn4 occupancy of open reading frame regions results in the recruitment of chromatin-modifying complexes but not the mediator complex. *EMBO Reports*, *4*(9), 872–876. https://doi.org/10.1038/sj.embor.embor931

Tripic, T., Deng, W., Cheng, Y., Zhang, Y., Vakoc, C. R., Gregory, G. D., … Blobel, G. A. (2009). SCL and associated proteins distinguish active from repressive GATA transcription factor complexes. *Blood*, *113*(10), 2191–2201. https://doi.org/10.1182/blood-2008-07-169417

Valenzuela, L., Aranda, C., & González, A. (2001). TOR modulates GCN4-dependent expression of genes turned on by nitrogen limitation. *Journal of Bacteriology*, *183*(7), 2331–2334. https://doi.org/10.1128/JB.183.7.2331-2334.2001

van Helden, J., André, B., & 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*(5), 827–842. https://doi.org/10.1006/jmbi.1998.1947

Venters, B. J., & Pugh, B. F. (2009). A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. *Genome Research*, *19*(3), 360–371. https://doi.org/10.1101/gr.084970.108

Wery, M., Descrimes, M., Vogt, N., Dallongeville, A. S., Gautheret, D., & Morillon, A. (2016). Nonsense-Mediated Decay Restricts LncRNA Levels in Yeast Unless Blocked by Double-Stranded RNA Structure. *Molecular Cell*, *61*(3), 379–392. https://doi.org/10.1016/j.molcel.2015.12.020

750 Wu, J., Zhang, N., Hayes, A., Panoutsopoulou, K., & Oliver, S. G. (2004). Global analysis of nutrient
751 control of gene expression in Saccharomyces cerevisiae during growth and starvation. *Proc Natl*
752 *Acad Sci USA*, *101*(9), 3148–3153.
753 Zambelli, F., Pesole, G., & Pavesi, G. (2013). Motif discovery and transcription factor binding sites
754 before and after the next-generation sequencing era. *Briefings in Bioinformatics*, *14*(2), 225–
755 237. https://doi.org/10.1093/bib/bbs016
756 Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., … Liu, X. S. (2008).
757 Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137.
758 https://doi.org/10.1186/gb-2008-9-9-r137
759

760 ## Figure Legends

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

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

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

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

774 **d.** Metagene view of the ChIP-Seq signal along the ATG +/- 600 bp region for the 1269 genes identified
775 as bound by Dal80 at the promoter (solid lines) and for the unbound genes (dashed lines), in untagged

776 (black) and Dal80-Myc[13] cells (blue). For each group of genes, normalized coverage (tag/nt) for each
777 gene was piled up, and the average signal was plotted. The shading surrounding each line denotes the
778 95% confidence interval.
779 **e.** Snapshot of ChIP-Seq signals along the *FUR4 locus*. Densities (tag/nt) are shown for the untagged
780 (black line) and Dal80-Myc[13] (blue line) strains. ORFs and the Dal80-Myc[13] ChIP-Seq peak are
781 represented by grey arrows and the red dashed line, respectively. The position (and orientation) of
782 each GATA site is represented by vertical segments above (sense GATA sites) or below (antisense GATA
783 sites) the *locus* line. The snapshot was produced using the VING software (Descrimes *et al.*, 2015).
784 **f**. Number of GATA sites in the promoter of Dal80-unbound (4522) and promoter-bound (1268) genes.
785 The analysis was performed using RSAT (Medina-Rivera *et al.*, 2015), across the -500 to -1 region
786 (relative to the ATG codon of the downstream ORF).
787 **Figure 2. Dal80 binds the intragenic region of a subset of genes.**
788 **a.** Identification of genes showing Dal80 intragenic binding.
789 **b.** Venn diagram showing the number of genes that are bound by Dal80 at the promoter and/or across
790 the ORF.
791 **c.** Metagene view of the ChIP-Seq signal along the ATG +/- 600 bp region for the 1125 genes identified
792 as bound by Dal80 at the promoter only ("P"), in untagged (black) and Dal80-Myc[13] cells (blue).
793 MetaSignal computation was as described in Fig. 1d. The shading surrounding each line denotes the
794 95% confidence interval.
795 **d.** Same as above for the 144 genes bound by Dal80 at the promoter and across the ORF ("P & O").
796 **e.** Same as above for the 45 genes bound by Dal80 across the ORF only ("O").
797 **f.** Same as above for the 4484 genes that are not bound by Dal80.
798 **Figure 3. Dal80 binding across gene bodies correlates with Dal80-sensitive, high expression levels.**
799 **a.** Scatter plot of densities (tag/nt, $\log_2$ scale) for genes in WT and *dal80Δ* cells grown in proline-
800 containing medium. Total RNA was extracted from exponentially growing biological replicates of 25T0b
801 (WT) and FV080 (*dal80Δ*). After rRNA depletion, strand-specific RNA-Seq libraries were constructed

802 and then sequenced. Tag densities were computed using uniquely mapped reads. Dal80-regulated
803 genes were identified on the basis of a min. 2-fold change (up or down) in the *dal80Δ* mutant relative
804 to the WT control, with a P <0.05 upon differential expression analysis using DESeq (Anders & Huber,
805 2010). Unaffected (n=5252), Dal80-repressed (n=232) and Dal80-activated (n=314) genes are shown
806 as grey, red and blue dots, respectively.

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

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

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

814 **e.** Cells of FV078 (wt *DAL80-MYC*[13]) strain were grown in proline as unique nitrogen source. Anti-Myc
815 ChIP was performed as described in Experimental procedures. Each histogram represents the average
816 (x10 000) of the value IP/IN (input or total chromatin) of at least 3 independent cultures on which two
817 IPs were performed. The associated error bars correspond to the standard error ($DAL5_{P1-P2}$, n=6;
818 $DAL5_{O3-O4}$, n=5; $GAP1_{P1-P2}$, n=4; $GAP1_{O9-O10}$, n=4; $GDH2_{P1-P2}$, n=4; $GDH2_{O5-O6}$, n=4; $MEP2_{P9-P10}$, n=11;
819 $MEP2_{O1-O2}$, n=7). Black histograms represent qPCR performed on promoter regions whereas grey ones
820 indicate qPCR data obtained within genes. Dotted line represents the highest no tag control levels.
821 Primers are described in Supp. Table 2: a, $DAL5_{P1-P2}$; b, $DAL5_{O3-O4}$; c, $GAP1_{P1-P2}$; d, $GAP1_{O9-O10}$; e, $GDH2_{P1-}$
822 $_{P2}$; f, $GDH2_{O5-O6}$; g, $MEP2_{P9-P10}$; h, $MEP2_{O1-O2}$.

823 **f.** ChIP analysis was conducted as in Panel E, with or without RNase treatment before
824 immunoprecipitation (n=3).

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

826 **a.** Schematic representation of the P_{MEP2}-*URA3 locus*, GATA sites (vertical segments above -sense GATA
827 sites- or below -antisense GATA sites- the *locus* line) and qPCR probe positions (black segments, *MEP2*
828 promoter; grey segments, *URA3* ORF).
829 **b.** Occupancy of Dal80-Myc[13] at the *MEP2* promoter and *URA3* ORF at the *MEP2* and *URA3 loci* in wild
830 type and P_{MEP2}-*URA3* strains. Cells of *DAL80-MYC[13]* wild type (FV078) and P_{MEP2}-*URA3* (FV808) strains
831 were cultivated at 29°C in the presence of proline (Pro) as unique nitrogen source. ChIP analysis was
832 conducted as in Fig. 4. The first histogram summarizes the data obtained by qPCR probing the promoter
833 of *MEP2* (a, $MEP2_{P9\text{-}P10}$), whereas the two following histograms are generated by qPCR performed with
834 oligos positioned in the ORF of *URA3* (b, $URA3_{O1\text{-}O2}$; c, $URA3_{O3\text{-}O4}$), either at the *URA3 locus* (in wild type
835 strain), or both at the *URA3* and *MEP2 loci*. Dotted line represents the highest no tag control levels.
836 The associated error bars correspond to the standard error (n=6).
837 **c.** Schematic representation of the *UGA4 locus*, GATA sites (vertical segments above -sense GATA sites-
838 or below -antisense GATA sites- the *locus* line) and qPCR probe positions (black segments, promoter;
839 grey segments, ORF).
840 **d.** Wild type and truncated (ΔLZ) versions of Dal80-Myc[13] used for the ChIP experiment at the *UGA4*
841 *locus*.
842 **e.** Occupancy of the *UGA4 locus* by Dal80-Myc[13]. Cells of FV078 (*DAL80-MYC[13]*; wt) and FV136
843 (*DAL80ΔLZ-MYC[13]*; ΔLZ) strains were grown at 29°C in the presence of proline (Pro) as unique nitrogen
844 source. ChIP analysis was conducted as described in Fig. 4. Black histograms represent qPCR performed
845 with promoter-specific primers (a, $UGA4_{P1\text{-}P2}$; b, $UGA4_{P9\text{-}P10}$) whereas grey ones indicate qPCR data
846 obtained with ORF-specific primers (c, $UGA4_{O1\text{-}O2}$; d, $UGA4_{O3\text{-}O4}$) at the *UGA4 locus*. The associated error
847 bars correspond to the standard error (n=6).
848 **Figure 6. Transcription elongation is necessary for Dal80 binding across *MEP2*.**
849 **a.** Schematic representation of the *MEP2 locus*, GATA sites (vertical segments above -sense GATA sites-
850 or below -antisense GATA sites- the *locus* line), and qPCR probe positions (black segments, promoter;
851 grey segments, ORF).

852 **b.** Occupancy of Dal80-Myc[13] at the *MEP2 locus* when transcription elongation is impaired. Wild type
853 (FV673) or *rpb1-1* (FV675) *DAL80-MYC[13]* cells were grown at 29°C in the presence of proline (Pro) as
854 unique nitrogen source, and shifted at 37°C for one hour at mid-log phase. ChIP analysis was conducted
855 as in Fig. 4. The first 2 black columns represent qPCR with 2 primer pairs within the *MEP2* promoter (a,
856 $MEP2_{P5-P6}$; b, $MEP2_{P13-P14}$) and the 2 following grey ones correspond to qPCR data obtained with 2
857 primer pairs within the *MEP2* ORF (c, $MEP2_{O11-O12}$ d, $MEP2_{O9-O10}$). The associated error bars correspond
858 to the standard error (n=6).
859 **c**. Coimmunoprecipitation of Dal80-Myc[13] with different phosphoforms of PolII. Total proteins were
860 extracted, immunoprecipitated with the indicated antibodies, and subjected to anti-Myc western blot
861 analysis.
862 **Figure 7. Model summarizing the correlations observed between Dal80 binding, Dal80 dependency**
863 **and expression levels.**
864

865 ## Supplementary Information

866 **Supplementary Table 1.xlsx : Strains used in this study.** (Nonet et al., 1987)

867 **Supplementary Table 2.xlsx : Primers used in this study.**

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

873 **Supplementary Table 4.xlsx - Related to Figure 2 : Dal80 binds the intragenic region of a subset of**
874 **genes.** Inventory of gene promoters and ORFs bound by Dal80 and their intersection.
875 **Supplementary Table 5.xlsx - Related to Figure 3 : Dal80 binding across gene bodies correlates with**
876 **Dal80-sensitive, high expression levels.** *Sheet a.* List of Dal80-regulated genes on proline. *Sheet b.*
877 Gene ontology term analysis of the 232 genes repressed by Dal80 on proline. *Sheet c.* Gene ontology
878 term analysis of the 314 genes activated by Dal80 on proline.
879 **Supplementary Figure 1.pdf - Related to Figure 1 : Genome-wide identification of Dal80-bound**
880 **promoters.**
881 **a.** Box-plot of the distance between the annotated TSS and ORF start site (translation initiation codon,
882 ATG) for protein-coding genes.
883 **b.** Snapshot of ChIP-Seq signals along a GATA-less *locus* (*ALD6*). Densities (tag/nt) are shown for the
884 untagged (black line) and Dal80-Myc^{13} (blue line) strains. ORFs and the Dal80-Myc^{13} ChIP-Seq peak are
885 represented by grey arrows and the red dashed line, respectively. The snapshot was produced using
886 the VING software (Descrimes et al., 2015).
887 **Supplementary Figure 2.pdf – Related to Figure 3 : Dal80 binding across gene bodies correlates with**
888 **Dal80-sensitive, high expression levels.**
889 **a.** Contingency table showing the number of Dal80-regulated or unaffected genes (defined upon RNA-
890 Seq analysis in WT and *dal80Δ* cells) among the Dal80-bound and unbound genes (identified upon
891 ChIP-Seq analysis in untagged and Dal80-Myc^{13} cells). Results that were experimentally observed or
892 that were expected in case of independence are indicated in bold and in brackets, respectively.
893 **b.** Metagene view of the ChIP-Seq signal along the 216 Dal80-regulated and -bound genes (solid lines),
894 and the 330 Dal80-regulated and -unbound genes (dashed lines), in the untagged (black) and Dal80-
895 Myc^{13} (blue) strains. The shading surrounding each line denotes the 95% confidence interval.
896 **c.** Density-plot of RNA-Seq signal (tag/nt, log2 scale) in the WT strain grown in proline-containing , for
897 genes of the "unbound" (blue, n=4484), "P" (red, n=1125) and P&O" (black, n=144) classes. Y-axis:
898 proportion of genes for each class. The highlighted areas correspond to the 75 (2%) and 170 (15%)

899 genes of the “unbound” and “P” classes, respectively, showing a signal higher than the median of the
900 “P&O” class. A box-plot representation of the same RNA-Seq signals is shown on the top of the density-
901 plot.
902 **d.** Same as c., highlighting the 949 (21%) and 632 (56%) genes of the “unbound” and “P” classes,
903 respectively, showing a signal higher than the first quartile value for the “P&O” class.
904 **Supplementary Figure 3.pdf – Related to Figure 4 : Dal80 binding across the ORF of well-**
905 **characterized NCR-sensitive genes – Control experiments.**
906 **a.** Cells of FV078 (wt *DAL80-MYC*[13]) strain were grown in glutamine (Gln) or proline (Pro) as unique
907 nitrogen sources. Anti-Myc ChIP was performed as described in Experimental procedures. Each
908 histogram represents the average (x10 000) of the value IP/IN (input or total chromatin) of at least 3
909 independent cultures on which two IPs were performed. The associated error bars correspond to the
910 standard error ($DAL5_{P1-P2}$, n=6; $DAL5_{O3-O4}$, n=5; $GAP1_{P1-P2}$, n=4; $GAP1_{O9-O10}$, n=4; $GDH2_{P1-P2}$, n=4; $GDH2_{O5-}$
911 $_{O6}$, n=4; $MEP2_{P9-P10}$, n=11; $MEP2_{O1-O2}$, n=7). Black histograms represent qPCR performed on promoter
912 regions whereas grey ones indicate qPCR data obtained within genes. Primers are described in Supp.
913 Table 2: $DAL5_{P1-P2}$, $DAL5_{O3-O4}$, $GAP1_{P1-P2}$, $GAP1_{O9-O10}$, $GDH2_{P1-P2}$, $GDH2_{O5-O6}$, $MEP2_{P9-P10}$, $MEP2_{O1-O2}$.
914 **b.** Gal4-Myc[13] binding at the *GAL1-10* promoter. Schematic representation of the *GAL1-10 locus*.
915 Position of the Gal4-binding sites (vertical segments across the *locus* line) and qPCR probes (black and
916 grey lettered segments) are indicated. Cells of FV437 (wt *GAL4-MYC*[13]) strain were grown in YP with
917 glucose or galactose as unique carbon sources. Anti-Myc ChIP was performed as described in
918 Experimental procedures. Each histogram represents the average (x10 000) of the value IP/IN (input
919 or total chromatin) of 2 independent cultures. The associated error bars correspond to the standard
920 error (n=2). Black histograms represent qPCR performed on promoter regions whereas grey ones
921 indicate qPCR data obtained within genes. Primers are described in Supp. Table 2.
922 **c.** Binding of Dal80-Myc[13] at the *MEP2 locus*, in untreated or RNase-treated cells. Schematic
923 representation of the *MEP2 locus*. Position of the GATA sites (vertical segments above -sense GATA
924 sites- or below -antisense GATA sites- the *locus* line) and qPCR probes (black and grey lettered

925 segments) are indicated. Cells of untagged (25T0b) or FV078 (*DAL80-MYC*[13]; wt) strains were grown at
926 29°C in the presence of proline (Pro) as unique nitrogen source. ChIP analysis was conducted with or
927 without RNase treatment before immunoprecipitation. The first 3 black columns represent qPCR with
928 3 primer pairs within the *MEP2* promoter $MEP2_{P5-P6}$; $MEP2_{P13-P14}$; $MEP2_{P9-P10}$) and the 3 following grey
929 ones correspond to qPCR data obtained with 3 primer pairs within the *MEP2* ORF ($MEP2_{O1-O2}$; $MEP2_{O11-}$
930 $_{O12}$ $MEP2_{O9-O10}$). Each histogram represents the average (x10 000) of the value IP/IN (input or total
931 chromatin) of independent cultures. The associated error bars correspond to the standard error (n=3).
932 **Supplementary Figure 4.pdf – Related to Figure 5 : Dal80 binding within gene bodies occurs**
933 **independently of intragenic GATA sites – Control experiments.**
934 **a.** Expression of the non-NCR-sensitive gene *URA3* under the control of P_{MEP2}. Cells of wild type (25T0b)
935 and P_{MEP2}-*URA3* (FV806) strains were cultivated at 29°C in the presence of proline (Pro) as unique
936 nitrogen source. The expression of *URA3* is normalized to the expression of the *TBP1* gene. Each
937 column corresponds to the mean value of data obtained for at least 2 independent cultures and the
938 error bars correspond to the standard error (n=2).
939 **b.** PolII occupancy at the *URA3* and P_{MEP2}-*URA3 loci*. Cells of *DAL80-MYC*[13] wild type (FV078) and P_{MEP2}-
940 *URA3* (FV808) strains were cultivated at 29°C in the presence of glutamine (Gln) or proline (Pro) as
941 unique nitrogen sources. Anti-polII (CTD4H8) ChIP was performed as described in Experimental
942 procedures. Each histogram represents the average (x10 000) of the value IP/IN (input or total
943 chromatin) of at least 2 independent cultures on which two IPs were performed. Histograms represent
944 qPCR performed with $URA3_{O1-O2}$ primers (n=2).
945 **c.** *UGA4* expression in a *DAL80ΔLZ* mutant strain. Cells of FV078 (*DAL80-MYC*[13]; wt) and FV136
946 (*DAL80ΔLZ-MYC*[13]; ΔLZ) strains were cultivated at 29°C in the presence of proline (Pro) as unique
947 nitrogen source. Total RNA was isolated and *TBP1*-normalized *UGA4* mRNA levels were quantified by
948 qRT-PCR using $UGA4_{O1-O2}$ primers. Histograms represent the averages of at least 2 independent
949 experiments and the associated error bars correspond to the standard error (n=3).

950 **d.** Occupancy of the *UGA4 locus* by PolII. Cells of FV078 (*DAL80-MYC*^{13}; wt) and FV136 (*DAL80ΔLZ-*
951 *MYC*^{13}; ΔLZ) strains were grown at 29°C in the presence of proline (Pro) as unique nitrogen source.
952 Anti-polII (CTD4H8) ChIP analysis was conducted as in Panel B using $UGA4_{O1-O2}$ primers (n=2).

953 **Supplementary Figure 5.pdf – Related to Figure 6 : Transcription elongation is necessary for Dal80**
954 **binding across *MEP2* – Control experiments.**

955 **a.** Effect of PolII elongation defects on *MEP2* expression. Wild type (FV673) or *rpb1-1* (FV675) *DAL80-*
956 *MYC*^{13} cells were grown at 29°C in the presence of glutamine (Gln) or proline (Pro) as unique nitrogen
957 sources, and shifted at 37°C for one hour at mid-log phase. Total RNA was isolated and *TBP1-*
958 normalized *MEP2* mRNA levels were quantified by qRT-PCR using $MEP2_{O9-O10}$ primers. Histograms
959 represent the average of at least 2 independent experiments and the associated error bars correspond
960 to the standard error (n=2).

961 **b.** PolII occupancy at the *MEP2 locus*. Wild type (FV673) or *rpb1-1* (FV675) *DAL80-MYC*^{13} cells were
962 grown at 29°C in the presence of glutamine (Gln) or proline (Pro) as unique nitrogen sources, and
963 shifted at 37°C for one hour at mid-log phase. ChIP analysis was conducted as in Supp. Fig. 4b using
964 $MEP2_{O9-O10}$ primers (n=4).

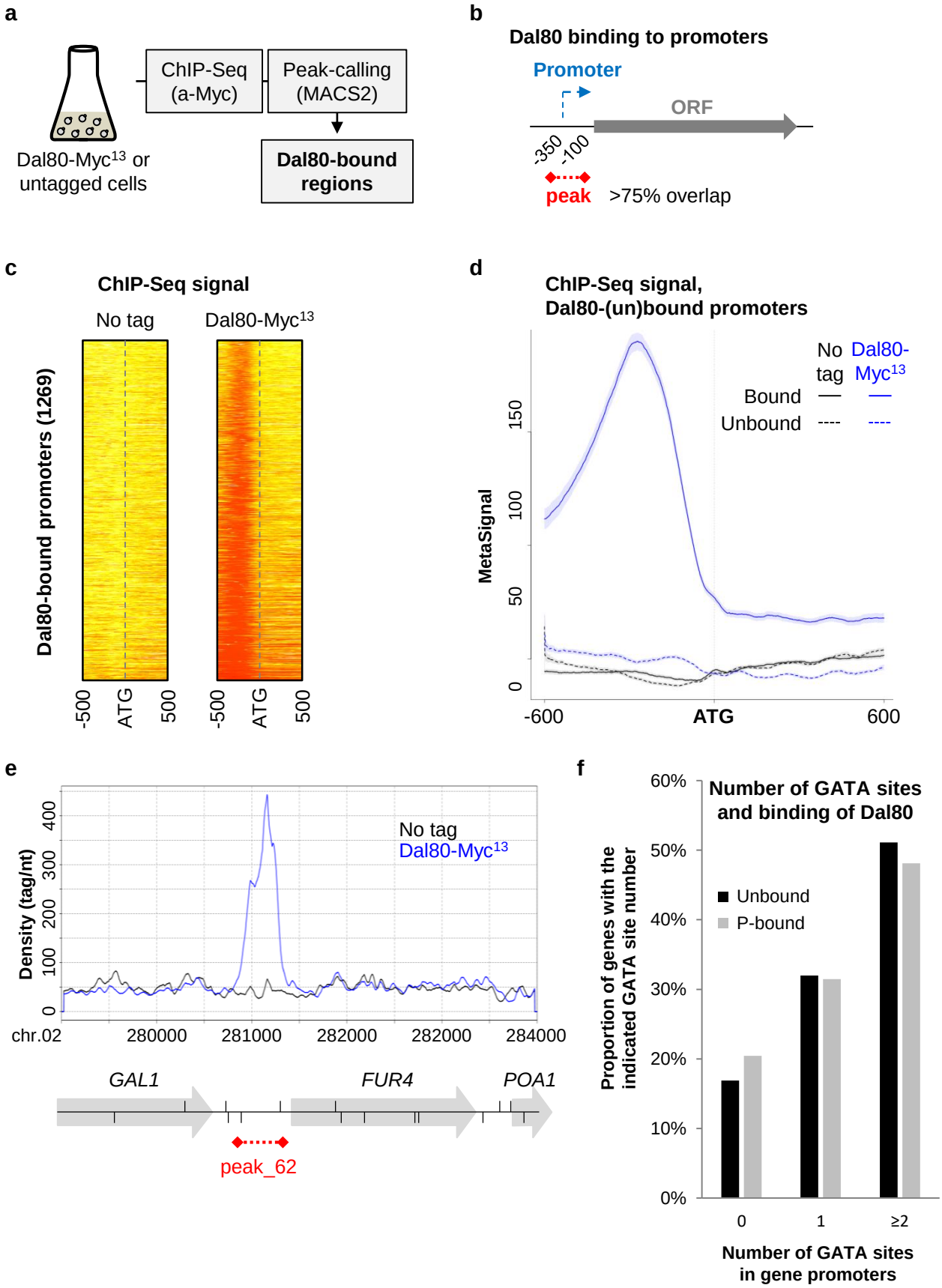

Figure 1. Genome-wide identification of Dal80-bound promoters. a. Overview of the ChIP-Seq analysis. **b.** Identification of Dal80-bound promoters. **c.** Heatmap view of the ChIP-Seq signal). **d.** Metagene view of the ChIP-Seq signal. **e.** Snapshot of ChIP-Seq signals along the *FUR4 locus*. **f.** Number of GATA sites in the promoter of Dal80-unbound (4522) and promoter-bound (1268) genes.

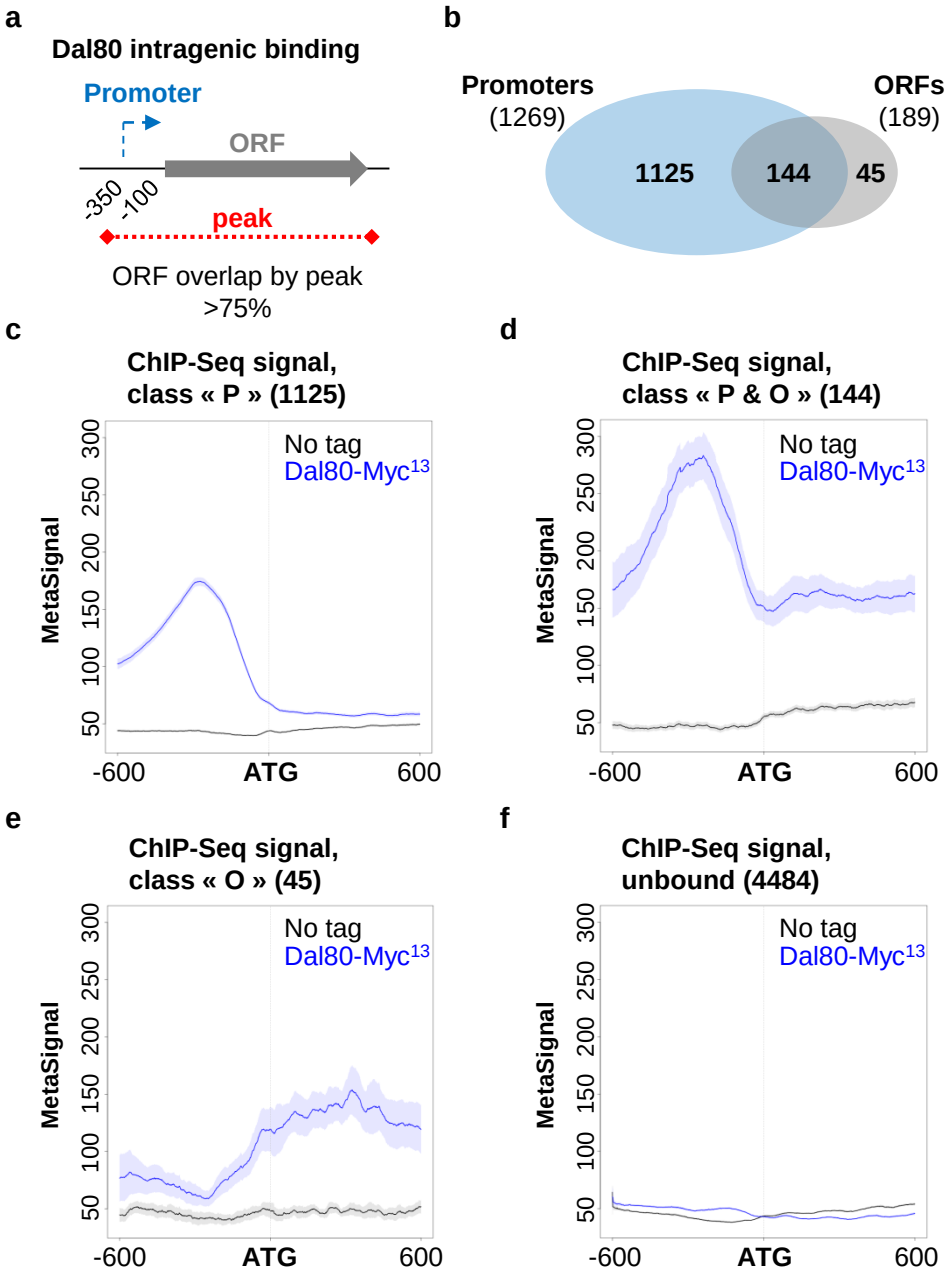

Figure 2. Dal80 binds the intragenic region of a subset of genes. a. Identification of genes showing Dal80 intragenic binding. **b.** Venn diagram showing the number of genes that are bound by Dal80 at the promoter and/or across the ORF. **c.** Metagene view of the ChIP-Seq signal along the ATG +/- 600 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). **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.

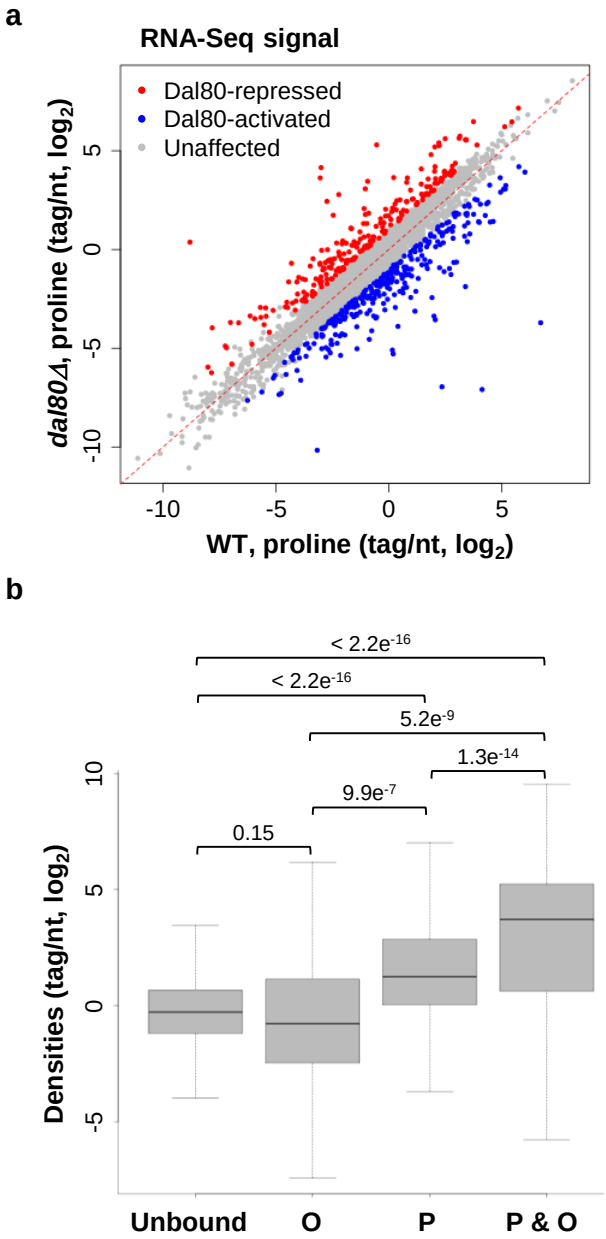

Figure 3. Dal80 binding across gene bodies correlates with Dal80-sensitive, high expression levels. a. Scatter plot of densities (tag/nt, $\log_2$ scale) for genes in WT and *dal80Δ* cells grown in proline-containing medium. **b.** Box-plot of densities (tag/nt, $\log_2$ scale) in the WT strain grown in proline-containing medium, for genes of the unbound, "O", "P" and "P & O" classes.

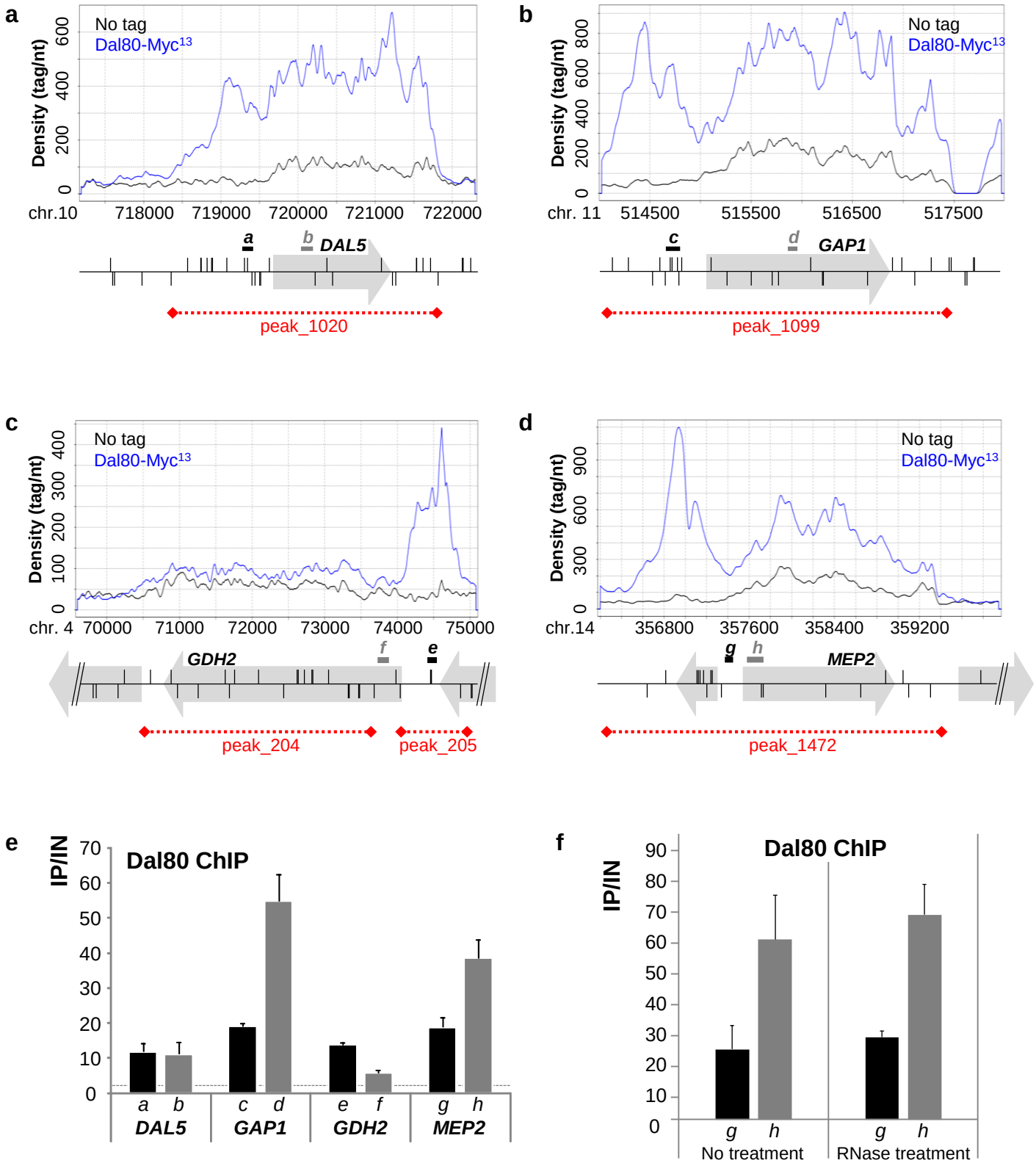

Figure 4. Dal80 binding across the ORF of well-characterized NCR-sensitive genes. a-d. Position of GATA sites (vertical segments above -sense GATA sites- or below -antisense GATA sites- the *locus* line) and qPCR probes (black and grey lettered segments) along 4 NCR-sensitive loci: *DAL5* (A), *GAP1* (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. **f.** ChIP analysis was conducted as in Panel E, with or without RNase treatment before immunoprecipitation.

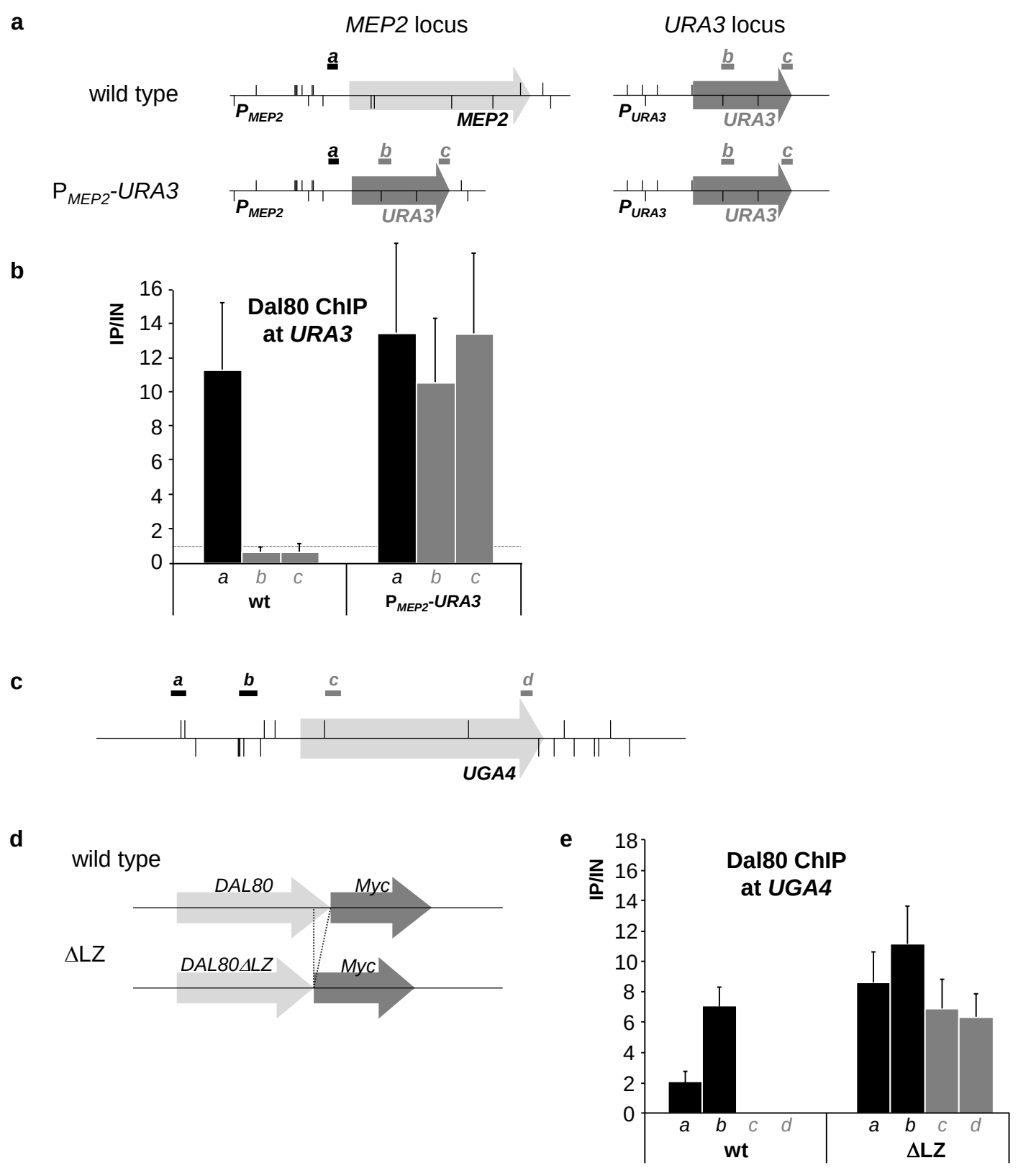

Figure 5. Dal80 binding within gene bodies occurs independently of intragenic GATA sites. a. Schematic representation of the P_{MEP2}-*URA3 locus*, GATA sites (vertical segments above -sense GATA sites- or below -antisense GATA sites- the *locus* line) and qPCR probe positions (black segments, *MEP2* 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. **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].

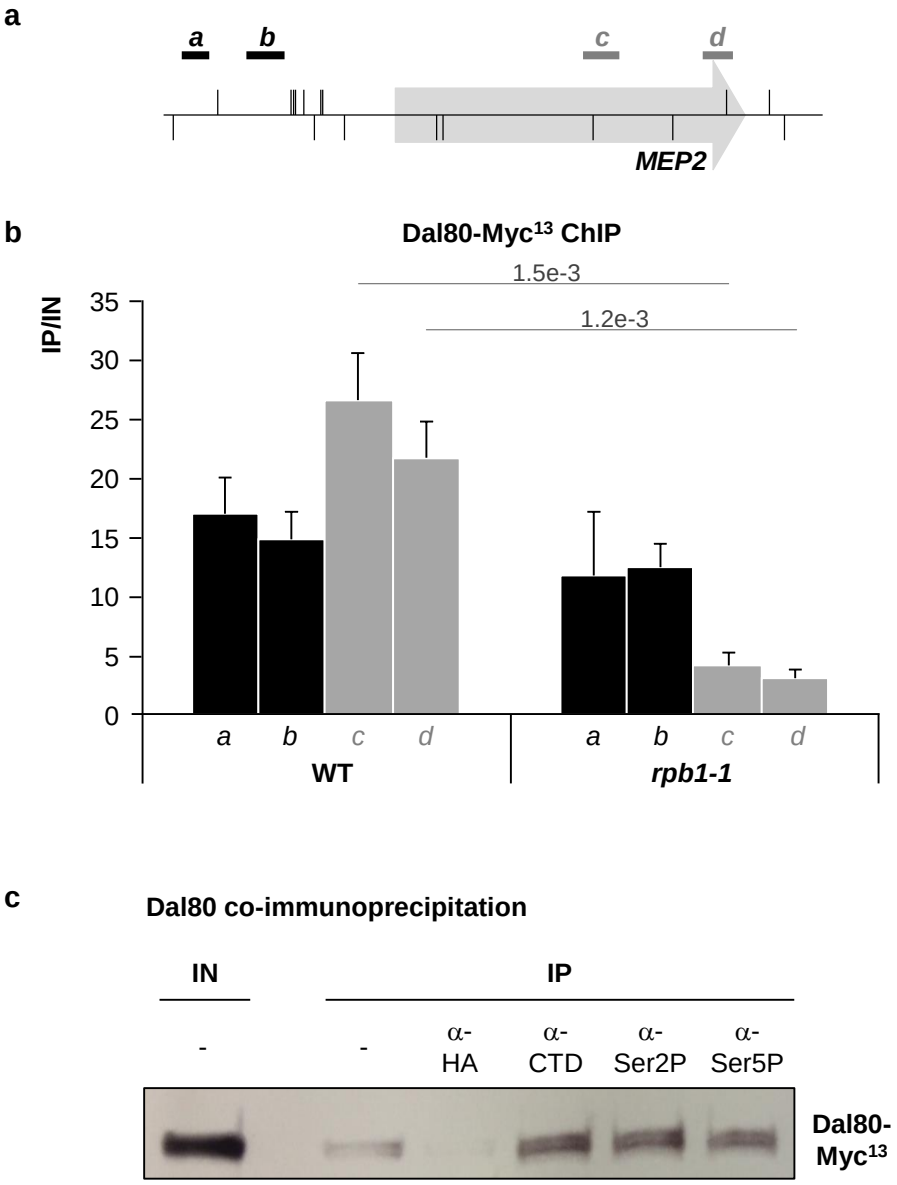

Figure 6. Transcription elongation is necessary for Dal80 binding across *MEP2*. a. Schematic representation of the *MEP2 locus*, GATA sites (vertical segments above -sense GATA sites- or below -antisense GATA sites- the *locus* line), and qPCR probe positions (black segments, promoter; grey segments, ORF). **b.** Occupancy of Dal80-Myc^{13} at the *MEP2 locus* when transcription elongation is impaired. **c.** Coimmunoprecipitation of Dal80-Myc^{13} with different phosphoforms of PolII.

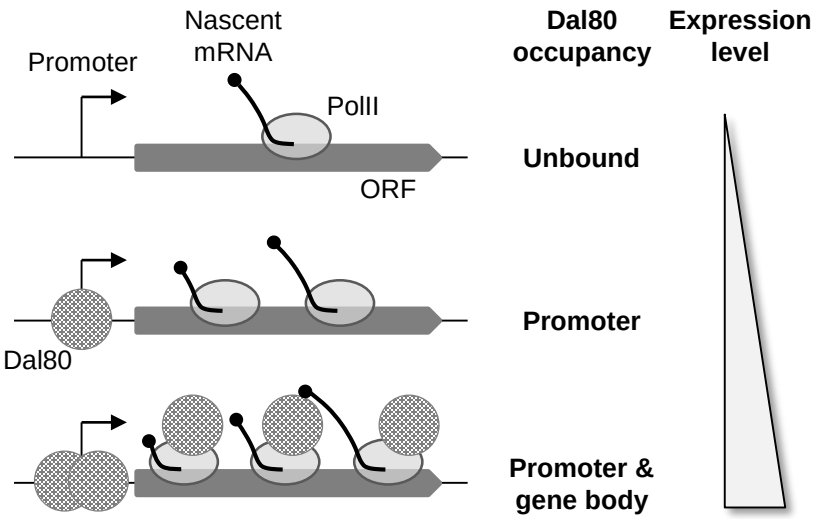

Figure 7. Model summarizing the correlations observed between Dal80 binding, Dal80 dependency and expression levels.