MAX-mutant small cell lung cancers exhibit impaired activities of MGA-dependent non-canonical polycomb repressive complex

Paula Llabata[1], Manuel Torres-Diz[2], Antonio Gomez[3], Laureano Tomas-Daza[4], Octavio A. Romero[1], Joaquim Grego-Bessa[2], Pere Llinas-Arias[5], Alfonso Valencia[6], Manel Esteller[5,7,8,9], Biola M. Javierre[3], Xiaoyang Zhang[10], Montse Sanchez-Cespedes[1,*]

[1]Cancer Genetics Group, Josep Carreras Leukaemia Research Institute (IJC), 08916 Badalona, Barcelona, Spain.

[2]Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), 08908 Hospitalet de Llobregat, Barcelona, Spain.

[3]Rheumatology Research Group, Vall d'Hebron Research Institute, 08035 Barcelona, Spain.

[4]3D Chromatin Organization Group, Josep Carreras Leukaemia Research Institute (IJC), 08916 Badalona, Barcelona, Spain.

[5]Cancer Epigenetics Group, Josep Carreras Leukaemia Research Institute (IJC) 08916 Badalona, Barcelona, Spain.

[6]Computational Biology Life Sciences Group, Barcelona Supercomputing Centre, 08034 Barcelona, Spain.

[7]Centro de Investigacion Biomedica en Red Cancer (CIBERONC), Madrid, Spain.

[8]Institucio Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain.

[9]Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona (UB), 08907 Barcelona, Spain.

[10]Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah UT 84112.

***Corresponding author:** E-mail: mscespedes@carrerasresearch.org

Abstract

The MYC axis is disrupted in cancer, predominantly through activation of the MYC-family oncogenes but also through inactivation of the MYC partner, MAX, or of the MAX partner, MGA. MGA and MAX are also members of the polycomb repressive complex, ncPRC1.6. Here, we use genetically modified MAX-deficient small cell lung cancer (SCLC) cells and carry out genome-wide and proteomics analyses to study the tumor-suppressor function of MAX. We find that MAX-mutant SCLCs have ASCL1 or NEUROD1 or combined ASCL1/NEUROD1 characteristics and lack MYC transcriptional activity. MAX restitution triggers pro-differentiation expression profiles that shift when MAX and oncogenic MYC are co-expressed. Although ncPRC1.6 can be formed, the lack of MAX restricts global MGA occupancy, selectively driving its recruitment towards E2F6 binding motifs. Conversely, MAX restitution enhances MGA occupancy to repress genes involved in different functions, including stem cell and DNA repair/replication. Collectively, these findings reveal that MAX-mutant SCLCs have either ASCL1 or NEUROD1 or combined characteristics and are MYC-independent and that exhibit deficient ncPRC1.6-mediated gene repression.

Significance

The MYC axis is commonly disrupted in cancer, mostly by activation of the MYC oncogenes, but also by genetic inactivation of MAX, the obligate partner of MYC, and of the MAX partner, MGA, both of which are members of the polycomb repressive complex, ncPRC1.6. While the oncogenic properties of the MYC family have been extensively studied, the characteristics of the MAX-deficient cells and the role of MGA in MAX-mutant cells remain unclear. In this study, we demonstrate that MAX-deficient SCLCs cells have either ASCL1 or NEUROD1 or combined characteristics. Furthermore, our data reveal that the lack of available MAX restricts MGA-occupancy in gene promoters and, although the ncPRC1.6 can still be formed, there is a deficient ncPRC1.6-mediated gene repression.

Introduction

Small cell lung cancer (SCLC), one of the most aggressive types of lung cancer, is commonly located centrally in the lung, and is thought to originate from the neuroendocrine cells of the lung epithelium (1-3). Mirroring the pattern of gene expression found in these cells, SCLCs have a high level of expression of neural-related transcripts (1,4-5). Recently, it has been shown that SCLCs comprise four subtypes, distinguished by the predominant expression of lineage-specific transcription factors. The most common of these are ASCL1 and NEUROD1, which target different gene sets for neuroendocrinal or neural functions6. The genetic profile of SCLCs includes the almost universal presence of inactivating alterations at *TP53*. Inactivating mutations at *RB1* and at *PTEN*, or activation of oncogenes, such as *MYC* and *PIK3CA*, are also characteristic of this type of lung cancer (7). With the advent of the novel sequencing technologies, novel genes, e.g., the *NOTCH* family, and genes coding for histone-modifiers, e.g., *CREBBP*, *EP300* and *MLL*, have been found to be altered in SCLCs (8-9).

The MYC axis is commonly disrupted in cancer, mostly by genetic activation of the *MYC* family of oncogenes. We reported that a subset of SCLCs features somatic and biallelic inactivation of *MAX*, a gene encoding for the obligate heterodimerization partner of the MYC family of proteins5. Recently, it has been shown that Max deletion increases growth and transformation in cells and dramatically accelerates SCLC progression in an Rb1/Trp53-deleted mouse model (10). Previously, germline mutations of *MAX* had been found in patients with hereditary pheochromocytomas, another neoplasia of neuroendocrine origin (11). The presence of gene alterations at MYC or MAX and of alterations of components of the SWI/SNF complex were found to be mutually exclusive, implying a functional connection between these pathways (5).

Although the genetic and molecular data strongly suggest that the loss of function of MAX contributes to the development of SCLC, it constitutes a conundrum in the understanding of MYC biological and molecular function, which depends on its dimerization with MAX12. Heterodimerization with MAX through the bHLH (basic helix-loop-helix) regions of both

proteins allows the recognition of the DNA sequences known as E-boxes. While MAX is the only partner of the MYC proteins and lacks a transactivation domain, it has a wide variety of other putative partners with which it could heterodimerize (i.e., MXD1, MXD2, MXD3, MXD4, MNT, and MGA) (12). It is now well established that the expression of the MYC target genes is controlled by the shift between activating MYC-MAX and repressive MAX-MXDs/MNT/MGA heterodimers that bind to the same canonical E-box consensus sequences in gene promoters (12). In this regard, the role of the MAX dimerization partner MGA may be especially significant because it is known to be genetically inactivated in a subset of cancers (5,13-14). MGA contains a bHLH domain, through which it binds MAX, and a T-box domain15 whose function is not understood. More recently, MAX and MGA have been shown to act as part of the Polycomb Repressive Complex 1 (PRC1), specifically the non-canonical PRC1, otherwise known as ncPRC1.6 (16-17).

The current study aimed to shed light on the tumor suppressor function of MAX and its relevance to SCLC development. Given that the MAX partner, MGA, is also inactivated in lung cancer and that both proteins are members of the polycomb repressive complex, ncPRC1.6, we wanted to define the possible role of MGA and of the ncPRC1.6 in cancer cells lacking MAX. Our findings demonstrate that MYC does not have any transcriptional function in MAX-deficient cells and that, in SCLC cells carrying MAX-inactivating mutations, there is a deficient ncPRC1.6-mediated gene repression which may contribute to cancer development.

Results

Small Cell Lung Cancer cells with MAX inactivation have ASCL1 or NEUROD1 or combined ASCL1/NEUROD1 characteristics. Previously, we identified four SCLC cell lines (COR-L95, H1417, Lu134 and Lu165) that lack MAX protein due to gene alterations (5). These cells have very low protein expression from the *MYC* family of genes (**Fig. 1A**). As mentioned above, SCLCs have been classified into four subtypes, based on the predominant expression of lineage-specific transcription factors (6,18). To molecularly characterize these MAX-deficient

cells further, we analyzed the gene expression profiles of a panel of SCLC cell lines (n=11), including *MYC*-, *MYCL1*-, and *MYCN*-amplified cells, using microarray gene expression analysis (***SI Appendix Fig. S1A***) and RNA-sequencing performed on the Lu134 and Lu165 cells lines and combined with RNA-sequencing data from the Cancer Cell Line Encyclopedia (n=50) (**Fig. 1B** and ***SI Appendix Fig. S1B***). The genetic status of MGA and of the SWI/SNF component, SMARCA4, was also annotated. As previously reported (6), most *MYC*-amplified cells clustered together and showed high levels of *NEUROD1* and of NEUROD1-targets, whereas most *MAX*-deficient, *MYCL1*- and *MYCN*-amplified cells, exhibit predominantly high expression levels of *ASCL1* and of ASCL1-targets. Some exceptions were the *MAX*-mutant, Lu134 cells and the *MYCL1*-amplified, HCC33 cells, with no detectable ASCL1 levels, by western-blot, while expressing NEUROD1 (**Fig. 1C**). On the other hand, the Lu165 express both ASCL1 and NEUROD1 (**Fig. 1B-C** and ***SI Appendix Fig. S1B***). Notably, the three *SMARCA4*-mutant cell lines in the study were low-*ASCL1*/low-*NEUROD1* but high-*YAP1* expressers (**Fig. 1B-C**). The high-*ASCL1* expressing group is enriched in various potassium channel transcripts (e.g. *KCNMB2*) and in other genes such as *GRP, ISL1*, and *RNF183*, among others, some of which are known targets of ASCL1 (**Fig. 1B**) (6). Likewise, the high-*NEUROD1* group features higher levels of NEUROD1 targets (e.g., *ANGPTL2*, *NEFM*, and *RGS10*). The comparative of the mRNA levels of selected ASCL1- and NEUROD1-targets among the *MAX*-mutant, *MYCN/MYCL1*-amplififed or *MYC*-amplified cell lines further evidenced this association (**Fig.1D**). Taken together, the results support the concept that most *MAX*-mutant SCLC cells show ASCL1 characteristics, although some can express only NEUROD1 or both NEUROD1 and ASCL1 factors.

Finally, we observed that, compared with the *MYC*-amplified, the *MAX*-mutant and the *MYCL1*- and *MYCN*-amplified cells express significant low mRNA levels of the MAX-binding partner, *MXD2*, while showing a tendency towards higher levels of *MXD3* and *MXD4* (***SI Appendix Fig. S1C***).

No occupancy of MYC at any DNA region in MAX-mutant cancer cells. As mentioned above, our current understanding of MYC biology posits that MYC's transcriptional activity depends on its dimerization with MAX (12). However, genome-wide evaluation of MYC recruitment to the DNA in cancer cells that lack MAX has never been attempted. In this context, natural MAX-knockout cells are an invaluable tool for determining whether the MYC family of proteins can bind DNA in the absence of MAX. Here, we used these cells to perform chromatin immunoprecipitation-sequencing (ChIP-seq) of MYC.

First, we used a doxycycline-inducible system to overexpress MYC (hereafter referred to as (hi)MYC cells) and the MYC and MAX proteins (hereafter (hi)MYC/MAX cells) in Lu134 cells (**Fig. 2A**). Next, we performed ChIP-seq of MYC and of MAX from these cells. We observed no occupancy of overexpressed MYC at any DNA region. In contrast, recruitment of MYC to the DNA could be readily detected in the (hi)MYC/MAX cells (**Fig. 2B-C**). These observations support the canonical view that dimerization with MAX is required to ensure the DNA-binding activities of MYC (19-22).

MAX-reconstituted *versus* MYC oncogenic-activated cells: conserved pattern of MAX bound to DNA but with an antagonistic gene expression profile. Traditional DNA-binding studies have shown that MAX is bound to the same DNA sequences regardless of its dimerization partner (21-23). To investigate this at a genome-wide scale and to evaluate the influence of the MYC and MAX protein levels (physiological-like *versus* supra-physiological) on the dynamics and distribution of MAX genomic occupancy, we rescued the expression of MAX in Lu134 and Lu165 cells by using a doxycycline-inducible system, because the stable expression of wild type MAX in these cell lines reduces cell growth (5). It has been reported that the supra-physiological levels of MAX shift the equilibrium from MAX heterodimers towards the formation of MAX/MAX homodimers (19-20), for which reason, we tested different doxycycline doses. Doxycycline concentrations of 10 and 5 ng/ml in the Lu134 and Lu165, respectively, produced physiological-like levels of MAX (hereafter, (lo)MAX cells), comparable to those in SCLC cell lines bearing wild type MAX (**Fig. 2D**; ***SI Appendix Fig.***

160 ***S2A***). Instead, 1000 ng/ml of doxycycline was used to produce supra-physiological levels of
161 MAX ((hi)MAX cells, from herein). The ectopic expression of MAX was verified to be
162 homogeneously widespread in all the cells (***SI Appendix Fig. S2B***).

163 Next, we performed ChIP-seq of MAX in different contexts: i) low-to-zero MYC and
164 physiological-like levels of MAX ((lo)MAX cells); ii) low-to-zero MYC and high MAX levels
165 ((hi)MAX cells); and iii) high MYC and physiological-like MAX levels ((hi)MYC/MAX cells).
166 The latter cells express supra-physiological levels of MYC, with a MYC/MAX ratio almost
167 equal to that in the *MYC*-amplified SCLC cell line, H82 (***SI Appendix Fig. S2C***), thereby
168 constituting a model of SCLC cells that have shifted from being MAX-deficient to being MYC-
169 activated. Similar to previous observations (22-25), about half of the regions that recruited
170 MAX were within or near gene promoters (± 3 Kb around the transcription start sites [TSSs])
171 (**Fig. 2E**). There were significantly fewer MAX-bound regions in the two (lo)MAX cells,
172 although most of the annotated promoters for single genes in the (lo)MAX cells were included
173 among those of the (hi)MAX cells (***SI Appendix Fig. S3A***). There was a significant overlap (*P*
174 = 0.0001, permutation test) between peaks in annotated promoters for single genes. This was
175 particularly pronounced between the two (hi)MAX cells (> 80%) but was less between the two
176 (lo)MAX cells (30-40%) (**Fig. 2F**). Global analysis of the ChIP-seq results and inspection of
177 multiple regions of the genome also showed a strong overlap between the promoter regions
178 bound by MYC and MAX in the (hi)MYC/MAX cells, reflecting their co-localization (**Fig. 2C**
179 and ***SI Appendix Fig. S3B***).

180 Our previous findings showed that the expression profile after restoring MAX in SCLC cells
181 was inversely correlated with that of the lungs of mice carrying activated Myc or Nmyc (5).
182 Supporting this, we observe here that the global changes in gene expression after rescuing
183 MAX, for (hi)MAX and (lo)MAX cells, were opposite to those after oncogenic activation of
184 MYC ((hi)MYC/MAX cells) (**Fig. 3A-B** and ***Dataset S1***). The overall changes were very
185 similar among (lo)MAX- and (hi)MAX-expressing cells (***SI Appendix Fig. S4***), which may be
186 consistent with the observation that although MAX can form homodimers, they are inhibited

from binding DNA (26). Despite these similarities, it is important to mention that the upregulation and downregulation was more marked in the (hi)MAX than in the (lo)MAX cells. Gene ontology (GO) enrichment analysis showed that MAX restitution activates the transcription of genes involved in differentiation while it represses genes involved in ribosome biogenesis, mitochondrial and cytoplasmic translation and energy metabolism, which are known to be activated by MYC in MYC-transformed cells (25). These processes were inversely regulated in the (hi)MYC/MAX cells (**Fig. 3C** and ***SI Appendix Fig. S5***). Gene set enrichment analysis (GSEA) showed a direct correlation of the (hi)MYC/MAX, and an inverse correlation of the (lo)MAX- and (hi)MAX-associated profiles, with the previously identified targets of Myc and Nmyc (**Fig. 3D**).

The interaction of MAX with MYC activates gene expression, but its interaction with other partners has repressive effects (12,20). Accordingly, we found stronger binding of MAX to repressed genes in (lo)MAX and (hi)MAX cells, and a predominant transcriptional repression among the targets of MAX in the (lo)MAX cells. In contrast, in the (hi)MYC/MAX cells, MAX bound more strongly in the activated genes and its recruitment to promoters was significantly associated with transactivation (**Fig. 3E** and ***SI Appendix Fig. S6A-C***).

Taken together, these results suggest that the profile of MAX recruitment to the genomic DNA is similar when MAX is overexpressed alone and when it is concomitantly expressed with oncogenic levels of MYC. However, the global patterns of gene expression are strongly shifted under the two circumstances, possibly as a result of the different activities of MAX arising from its binding to distinct partners.

The absence of MAX does not affect the formation of the ncPRC1.6. As mentioned above, the MYC-MAX and MXDs/MNT/MGA-MAX complexes have opposite or antagonistic functions in transcriptional regulation, with MAX being required for DNA binding by all the factors in the network. MAX also acts as part of the non-canonical Polycomb Repressive Complex 1 (ncPRC1), specifically ncPRC1.6, which includes the following set of proteins:

E2F6, L3MBTL2, MGA, PCGF6, RING1A, RING1B, RYBP, TFDP1, YAF2, and WDR5 (**Fig. 4A**) (17, 27-28). Given this, we examined whether the presence or absence of MAX affects the formation or composition of ncPRC1.6. We profiled the binding of MGA by immunoprecipitation of endogenous MGA then carried out mass spectrometry-based proteomic analysis in the Mock, Lu134 and Lu165 cells and after restitution of MAX ((lo)MAX cells). MGA was found to associate with all components regardless of the presence of MAX (**Fig. 4B**). We confirmed these results by immunoprecipitating MGA, followed by immunoblotting (**Fig. 4C**). Similar results were obtained after immunoprecipitating E2F6, followed by immunoblotting (***SI Appendix Fig. S7A***). Our results demonstrate that ncPRC1.6 forms regardless of the presence or absence of MAX. In addition to the known protein constituents of ncPRC1.6, the mass spectrometry proteomics screening identified other proteins bound to MGA (***Dataset S2).*** Additional studies are needed to assess their interaction with MGA and with the ncPRC1.6 and their functional implications.

As a type of PRC1, ncPRC1 catalyzes the monoubiquitination of histone H2A at lysine 119 (H2AK119ub1) through the heterodimeric E3 ligase RING1B/PCGF1–6, and thereby contributes to chromatin compaction and transcriptional silencing (28). We did not find any changes in the global levels of H2AK119ub1 upon restitution of MAX or in the (hi)MYC/MAX cells ***(SI Appendix Fig. S7B***).

MAX reconstitution enhances the recruitment of MGA to the DNA and represses genes with cell division- and germ cell-related functions. Since ncPRC1.6 formation is independent of MAX, we next investigated how the availability of MAX affects the DNA-binding activities of MGA. We performed ChIP-seq of MGA in the various Lu134 and Lu165 cell models (i.e., Mock, (hi)MAX, (lo)MAX and (hi)MYC/MAX cells). Similar to previous reports (29,30), our analysis confirmed that MGA was bound in close proximity to TSS. Further, we observed that MGA was recruited to the DNA in the Mock cells, although the rescue of MAX expression leads to a global gain of MGA occupancy, including the recruitment of MGA to additional gene promoters (**Fig. 5A**). It was also observed that the promoters bound by MGA, in any of the cell

models, are targets of MAX, as is evident in the (hi)MAX cells, in which MAX is bound to more than 90% of the promoter targets of MGA (**Fig. 5B**). We found that the restitution of MAX drove moderate changes in gene expression among the MGA-associated promoters in the (hi)MAX and (lo)MAX cells, predominantly transcriptional repression (**Fig. 5C** and ***SI Appendix Fig. S8***). In contrast, in the (hi)MYC/MAX cells, the targets of MGA showed changes in gene expression consisting on both transcriptional activation and repression.

The observations above imply that MGA can bind some promoter regions in the absence of MAX while other promoters recruit MGA only when MAX is restored, thus constituting MAX-independent and MAX-dependent targets of MGA, respectively. Here, we classified these, respectively, as promoters with type 1 and type 2 binding sites (hereafter, BS1 and BS2). We generated lists of BS1- and BS2-associated promoters according to the criteria that BS1 were gene promoters that recruited MGA in either of the Mock cell lines, and that BS2 were gene promoters that recruited MGA in the (hi)MAX cells, unless they had already been classified as BS1. The BS1 promoters were less abundant than the BS2 promoters in both cell lines (**Fig. 5D**). We selected approximately the top 10% BS1 and BS2 promoters with the highest intensity of binding in each group for detailed analysis (***Dataset S3***). GO term enrichment analysis revealed that the BS2-associated genes were enriched in cell differentiation, apoptosis and metabolic-related features, whereas the BS1-associated genes were related to transcription and DNA replication and repair processes (***SI Appendix Fig. S9***). BS1 and BS2, in (lo)MAX and (hi)MAX cells, were predominantly associated with transcriptional repression; only a few BS2-associated genes were upregulated, whereas in (hi)MYC/MAX cells, there was upregulation and downregulation associated with the recruitment of MGA to both types of promoters (**Fig. 5E**). The presence of gene activation and repression among the targets of MGA in the (hi)MYC/MAX could be due, at least in part, to competition between MGA and MYC to bind to MAX. Consistent with this, we observe a mutually exclusive pattern of MGA and MYC intensity of binding to the DNA in (hi)MYC/MAX cells (**Fig. 5F**).

Most targets of MGA that became activated in the (hi)MYC/MAX cells were found repressed in the (lo)MAX and (hi)MAX cells (Group I) (**Fig. 5F**). However, the opposite was not true, since the genes repressed in the (hi)MYC/MAX cells barely changed in the (lo)MAX and (hi)MAX cells (group II). Group II included important tumor suppressor genes, e.g., *KEAP1* and *FANCA*, as well as components of ncPRC1.6, such as *RYBP, E2F6*, and *MGA* itself. There was a third group (III), comprising those genes that were repressed in the three cell models, such as TAF7L, GLS2 and HLTF, which were involved in germ cell-related processes (**Fig.5E-*F*** and ***Dataset S3***). This is in keeping with the findings in mouse pluripotent stem cells that the heterodimeric MGA/MAX is required to repress germ cell-specific genes (29). The level of these transcripts was found to be higher in lung cancer cells carrying *MGA* inactivation (***SI Appendix Fig. S10***). Further, the generation of knockouts for MGA in the A549 and H23 lung cancer cell lines, which are wild type for MGA and for MAX5, increased the levels of these transcripts, specially of *STAG3*, but not of *TAF7L* (**Fig. 6A-B**). A ChIP-sequencing analysis of the A549 cells, of E2F6, MGA and MAXs confirmed their recruitment to the *STAG3, GLS2* and *HLTF*, but not to *TAF7L,* promoters (***SI Appendix Fig. S6C***). Instead, the depletion of *MGA* in the Lu134 and Lu165 cells rendered no significant changes in the expression of these genes. All the above demonstrate that these transcripts are repressed by MGA, through the ncPRC1.6 complex.

MAX restitution shifts the DNA-binding profile of MGA from E2F motifs to E-boxes. As part of ncPRC1.6, MGA also interacts with heterodimeric E2F6/DP1/2 proteins, which bind DNA in a sequence-specific manner (29-30). Taking this into consideration, we studied the dynamics of the recruitment of MGA to E2F motifs (GCGGGA) or to E-boxes (CACGTG) depending on whether ectopic MAX is absent or present with or without oncogenic MYC. First, we determined the preferential binding of MYC, MAX, and MGA to these motifs, under the different conditions. As expected, MYC was bound almost exclusively to E-boxes, whereas MAX and MGA could be recruited to E-boxes and E2F motifs (**Fig. 7A-B**). A shift of MGA positioning from E2F motifs to E-boxes was observed in parallel with the restitution of MAX. This effect was stronger in the (hi)MYC/MAX cells (**Fig. 7B**).

Studying the distribution of the E2F motifs and E-boxes among the top 10% BS1 and BS2 promoters, we found that > 90% contained either one or both motifs. There was a widespread presence of the E2F motif, which was found in at least 80% of the BS1 and BS2 promoters. E-boxes were significantly over-represented throughout the BS2, and E2F sites alone were over-represented in BS1 (**Fig. 7C**). The over-representation of E2F motifs in BS1 indicates a preference for E2F6/DP1 binding. In this regard, it has been reported (30) that MGA is essential for recruiting ncPRC1.6 to its targets genes and that it executes its function through two different mechanisms: as a scaffold, that is independent of the bHLH domain but dependent on E2F6; and by sequence-specific bHLH-binding that is independent of E2F6. Many of the promoters from our BS1 lists match those regions previously found to recruit MGA in a bHLH-independent/E2F6-dependent manner (e.g., *RFC1, PHF20, SPOP,* and *RPA2*), whereas the BS2 lists include promoters that were found to recruit MGA in a bHLH-dependent/E2F6-independent manner (e.g., *CDIP, ZFR*) (**Fig. 5F, Fig. 7D** and ***Dataset S3***). Our current findings, combined with those of previous reports, indicate that, in the cancer cells that lack MAX, MGA acts essentially as a scaffold to recruit ncPRC1.6 to E2F6/DP1-dependent binding sites. The precise transcriptional interactions between E2F6/DP1 and MGA/MAX on BS1 are not yet fully understood, since both have been associated with transcriptional repression. Different affinities for the different promoters may account for the diverse and complex regulation of gene expression observed in these distinct genetic backgrounds and competition between MGA and MYC to heterodimerize with MAX, may also play a role in the case of the (hi)MYC/MAX cells. An example of these dynamics is the *AK2* gene, which is repressed in (hi)MAX and (lo)MAX cells, but activated in (hi)MYC/MAX cells. The *AK2* promoter has an E2F motif and an E-box that are distant enough to produce independent peaks in the IGV. In the absence of MYC ((lo)MAX and (hi)MAX cells) the MAX/MGA heterodimer is found only in the E2F motif, whereas, after MYC overexpression ((hi)MYC/MAX cells), the MAX/MYC heterodimer is bound to the E-boxes and the MAX/MGA is bound to the E2F motifs (**Fig. 7D**).

Oncogenic MYC reduces the level of *ASCL1* and of ASCL1-related transcripts and promotes NEUROD1 characteristics. Here, we showed that most *MAX*-deficient SCLCs expressed high levels of *ASCL1*. However, one of the two MAX-deficient cell lines studied in deep here, Lu134, predominantly express NEUROD1 while the other one, Lu165, express ASCL1 and NEUROD1 (**Fig. 1B-C**). We found that when MAX was overexpressed concomitantly with oncogenic levels of MYC ((hi)MYC/MAX cells) the mRNA levels of *ASCL1* were decreased by half in both cell lines, while the levels of *NEUROD1* were reduced in the Lu134 but not in the Lu165 cells (***Datase S1***). The effects on *ASCL1* and *NEUROD1* were not mediated by direct transcriptional regulation of MAX, MYC or MGA, since we were unable to detect MAX, MYC or MGA directly occupying the promoters of *ASCL1* or *NEUROD1* (**Fig. 6E**). Concomitant with the reduction in *ASCL1* there was prominent downregulation of most ASCL1 targets and upregulation of some NEUROD1 targets (31) in (hi)MYC/MAX cells compared with the Mock, MAX-deficient cells (**Fig. 7F** and ***SI Appendix Fig. S11***). Since *NEUROD1* was not upregulated, its relative increase over *ASCL1* abundance may underlie the observed shift from ASCL1 to NEUROD1 characteristics. These findings suggest that the transformation from a MAX-deficient to a MYC-oncogenic SCLC also shifts their dependency on these two transcription factors that are associated with the development of the neural lineage.

Discussion

We have shown that most SCLC cells bearing *MAX*-gene inactivation have ASCL1 characteristics, as is also the case for *MYCN* and *MYCL1*-amplified SCLC cells, suggesting that they have a degree of similarity or common origin. This is in contrast to the *MYC*-amplified SCLCs, which, as previously reported, have NEUROD1 characteristics (6,31-32). It is interesting that one MAX-deficient cell line used here express both the ASCL1 and the NEUROD1 factors, although the predominant profile was that of ASCL1. Our data also show that the expression of oncogenic MYC represses ASCL1 and triggers, to some extent, a NEUROD1 expression profile without upregulating *NEUROD1*. This suggests that the two

346 neurogenic transcription factors compete to establish a predominant genetic program. We did
347 not observe recruitment of MYC or MAX to the ASCL1 or NEUROD1 promoters, indicating
348 that other targets of MYC/MAX mediate the shift from ASCL1 to NEUROD1 characteristics in
349 SCLC cells.

350 The MYC family of proteins were barely expressed in the SCLC cells with genetically
351 inactivated MAX. A recent study showed MAX deficiency to have a profound effect on MYC
352 stability in both normal and premalignant settings (22), which supports our observations. Here,
353 we also observed a lack of MYC recruitment to the genomic DNA, even after ectopic
354 overexpression of MYC, implying that the transcriptional activity of the MYC proteins does not
355 play a role in the tumorigenesis of MAX-deficient cells. The concomitant ectopic expression of
356 MAX with oncogenic level of MYC restored the ability of MYC to bind DNA, producing a
357 gene expression profile compatible with that of *MYC*-amplified cancer cells. While the
358 requirement of MAX for the transcriptional activities of the MYC family of proteins has been
359 known for a long time (20-23,) even in a recent work using wide-genome screenings (22), our
360 current study is the first to demonstrate this in naturally MAX-deficient cancer cells. The targets
361 of the heterodimer MYC/MAX in these cells were associated with transactivation whereas the
362 targets of MAX in the MAX-restituted cells were mostly downregulated. Competition between
363 MYC and MXD1-4/MNT/MGA proteins for binding to MAX and to the same E-boxes is
364 known (21-23) and is evidence that these transcriptional regulators act as functional antagonists.
365 In this regard, we recently reported that the overexpression of MGA in lung cancer cells
366 represses the targets of MYC, consistent with the idea that they possess competitive and
367 antagonistic functions (33). Our new results fully support this view and also suggest that, since
368 MAX serves as a network edge, the genetic inactivation of MAX may contribute to the
369 development of cancer by preventing the pro-differentiation transcriptional regulation exerted
370 by its partnering with MXD/MNT/MGA.

371 Unlike MYC, MAX is an abundant and stable protein that is expressed in proliferating and
372 resting normal cells (20). However, we previously showed that MAX expression can be

regulated by corticoids (5). Others have shown that Max expression is transiently attenuated in germ cells undergoing meiosis *in vivo* and that the knockdown of Max in embryonic stem cells activates the expression of germ cell-related genes (34), indicating that the levels of MAX are regulated in some specific physiological processes. The role of MAX in regulating germ cell- and meiosis-related genes is dependent on ncPRC1.6, a PRC1 that includes MAX and MGA (27-29), the latter of which is also genetically inactivated in cancer (5, 13-14). In this study, we found that the lack of MAX does not prevent the formation of the complex but restricts the recruitment of MGA to the DNA. We defined as BS1 those sites within promoters that can recruit MGA in a MAX-independent manner, and as BS2 those that recruit MGA only after MAX-restitution. We also found that BS1 are enriched in E2F motifs, compared with BS2, which have more E-boxes. Remarkably, BS1-associated promoters were coincident with those regions previously reported to recruit MGA in a bHLH-independent/E2F6-dependent manner in which MGA acts as a scaffold, whereas the BS2-associated promoters matched those that recruit MGA in a bHLH-dependent/E2F6-independent manner (30). This leads us to postulate that, in cancer cells lacking MAX, MGA has a preeminent scaffolding function. In this scenario, the regulation of gene expression from ncPRC1.6 would be directed by the E2F6/DP1 module. Interestingly, several years ago we found that the gene coding for DP1, *TFDP1*, is strongly amplified in a small subset of lung tumors, leading to high levels of DP1 protein (35). Such levels of DP1 could have oncogenic potential, promoting the activities of the E2F6/DP1 module within ncPRC1.6. Our data confirms and extends previous knowledge about the various scenarios in which the competition for available MAX is important for different cell physiological processes, including cancer development (**Fig. 7G**).

In conclusion, we have demonstrated that most MAX-mutant SCLCs have ASCL1-like characteristics and are MYC-independent, and that exhibit deficient ncPRC1.6-mediated gene repression.

Material and methods

Lung cancer cell lines. Cell lines were obtained from the American Type Culture Collection (ATCC, Rockville, MD, USA), grown under recommended conditions and maintained at 37ºC in a humidified atmosphere. All cells tested negative for mycoplasma infection. The antibodies used are described in ***SI Appendix Table S1A***.

Western-blot, immunofluorescences and quantitative RT-PCRs. Antibodies and primers sequences, in ***SI Appendix Table S1B***. Detailed information about the methodologies are included in the SI Appendix.

Construction of expression vectors and infections. The complete MAX transcript (NM_145112.2) had been previously cloned[5]. Complete MYC (NM_002467.6) cDNA was PCR-amplified, from a retrotranscribed human RNA pool (Agilent Technologies, Santa Clara, CA, USA) using Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Waltham, MA, USA) following standard protocols (SI Appendix). The primers used are indicated in ***SI Appendix Table S1B***.

Microarray global gene expression analysis. We followed previously described procedures (5). The analysis was undertaken at the Genomics Unit of the Center for Genomic Regulation (CRG, Barcelona, Spain) (SI Appendix). Expression data were analyzed using the R statistical language (R Core Team (2014). URL http://www.R-project.org/). Raw data were extracted, and the background was corrected and normalized using the quantile algorithm available in Bioconductor's limma package (36). Normalized expression values were plotted with Bioconductor's ggplots and Complexheatmap packages. The ASCL1 and NEUROD1-target lists were elaborated selecting common genes occupied by ASCL1 or NEUROD1 in at least two cell lines, from the previous publication[6].

Chromatin immunoprecipitation (ChIP) sequencing. For ChIP assays, cells were fixed with 1% formaldehyde methanol-free (Thermo Scientific) for 10 min at room temperature and were then quenched by 125 mmol/L glycine for 15 min at room temperature, washed with ice-cold PBS twice and centrifuged at 200 *g*, at 4°C for 5 min. For each ChIP reaction, 60 µL of Magna ChIP™ Protein A+G Magnetic Beads (Merck, Millipore) was used according to the manufacturer's protocol. Detailed information about the methodology are included in the SI Appendix. At least two independent ChIP experiments were performed. Immunoprecipitated chromatin was deep-sequenced at the Centre for Genomic Regulation (CRG, Barcelona, Spain) using the Illumina HiSeq 2500 system.

ChIP sequencing data analysis. Reads were aligned to the human reference genome hg38, using Bowtie v1.2.2 with default parameters without allowing for multi-mapping (–m 1) (37). PCR duplicates were removed using PICARD (http://broadinstitute.github.io/picard/). Ambiguous mapping reads were discarded. Peaks were called using MACS2 v2.1.1 (38). To avoid false positives, peaks were discarded if they were present in the ChIP-seq of MAX in the MAX-deficient cells of the respective SCLC cells. Genomic peak annotation was performed with the ChIPpeakAnno v3.15 R package, considering the region of ± 3 Kb around the TSS as the promoter (39). Unless otherwise specified, all analyses considered the peaks overlapping with promoter regions. Peak lists were then transformed to gene target lists. Permutation tests (10,000 permutations) were performed to determine associations by overlap between region sets A and B, creating random regions throughout the genome using the Bioconductor package regioneR (40).

Bedgraph files were generated using the function makeUCSCfile from HOMER with default parameters normalizing for differences in sample library size, and BigWig files were generated using the function bedGraphToBigWig from UCSC. Heatmaps and intensity plots were performed using the functions computeMatrix, in a window of ±3kb center in the TSS, followed by plotHeatmap from deepTools (41). To homogenize the scale of all heatmaps and intensity plots, signal intensity was scaled to 0–1 by applying the formula (X–P05)/(P98–P05) to each matrix generated by computeMatrix.

Motif enrichment analyses were performed using HOMER motif discovery software (42). For annotated ChIP-seq peaks, a window of ± 100 bp around the peak center was applied. Values of $P < 0.01$ were taken to define a motif as being significantly enriched.

RNA sequencing. RNA sequencing was carried out at the Spanish National Genome Analysis Center (CNAG, Barcelona, Spain). About 2500 ng of total RNA from SCLC cell lines were used. RNA Integrity values ranged from 9.0 to 10.0 when examined by a BioA RNA Nano kit (Agilent). RNA-seq paired-end reads were mapped against the human reference genome (GRCh38) using STAR version 2.5.3a with ENCODE parameters for long RNA. DEseq2 was used to normalize counts. Annotated genes (gencode v27) were quantified using RSEM version 1.3.0 with default parameters. The RNA-seq report is provided in ***SI Appendix Table S1C***.

To generate the lists of upregulated and downregulated transcripts for each condition we chose the following criteria: i) transcripts induced or repressed, as indicated in each case, under each condition with respect to the Mock cell line, and ii) statistical significance (see below). The genes are listed in ***Dataset S1***.

The listed genes were subjected to several analyses, such as gene ontology (GO) functionalities (https://david.ncifcrf.gov) or gene set enrichment analysis (GSEA), using the indicated gene expression signatures (ranked by the n-fold values of change) as the gene set.

Immunoprecipitation (IP) and mass spectrometry. For IP, we used previously described protocols (33). Details of the antibodies in ***SI Appendix Table S1A***. 10 million cells were lysed with NP40 lysis buffer and sonicated with a tip ultrasonic homogenizer. Detailed information about the methodologies are included in the ***SI Appendix Methods***. For mass spectrometry assays, 100 million cells were grown in suspension and harvested by centrifugation. The cell pellet was resuspended in Net2 buffer (50 mmol/L Tris-HCl, 200 mmol/L NaCl, 0.1% Triton X-100, 5 mmol/L $MgCl_2$, 1 mmol/L EDTA) and incubated while rotating at 4ºC for 1 h. Cell lysate was sonicated and centrifuged at 2,500 *g*, at 4ºC for 3 min. Supernatant was collected and mixed with 20 µg of primary MGA antibody or IgG and incubated while rotating overnight at 4ºC. Details about the mass spectrometry assays can be found at the SI Appendix. Eluted and desalted peptides were resuspended in 10 µL 0.1% formic acid and loaded into the Orbitrap Velos Pro using the 'STD-VL-DDA-60min-T20-CID-IT' method. Peptides were analyzed with the Proteome Discoverer v1.4. with the 'STD-PWF-MASCOT-ANY-IT-DECOY' workflow. Peptides were filtered at 5% FDR.

We analyzed the results based on the enrichments of peptide precipitated by anti-MGA antibodies versus IgG controls (not found in the IgG control) and that were common to the two cell models. MAX, which is known to be the canonical binding partner of MGA (15), and the MGA protein itself are among the most significantly enriched proteins in this immunoprecipitation (***Dataset S2***).

Generation of MGA-depleted cell lines. The A549, H23, Lu134 or Lu165 cells were infected with lentivirus using plasmids (LentiCRISPR v2 (#52961, Addgene) expressing mammalian-codon optimized Cas9 and different sgRNAs targeting the coding region of human MGA or a non-target sequence, as negative control. The sequences of the oligonucleotides are included in the ***SI Appendix Table S1B***. Puromycin selection (2 µg/ml) was carried out 48 hours after infection for 3-4 days. Cell clones were further analyzed by Western blot.

Statistical analysis. Data were analyzed using a two-tailed Student's unpaired-samples t test or by Pearson's chi-square test. Group differences were presented as means and standard deviations. Differences were considered statistically significant for any value of $P < 0.05$.

Accession codes. Microarray gene expression data is available in the Gene Expression Omnibus (GEO) under accession codes GSE144457. The ChIP-seq and RNA-seq data obtained in this study have been uploaded to the SRA (NCBI), under accession number BioProject: PRJNA608275. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD017658.

References

1.-Y. Maeda, V. Davé, J.A, Whitsett. Transcriptional control of lung morphogenesis. *Physiol. Rev.* **87**, 219–244 (2007).

2.-K.D. Sutherland et al., Cell of origin of small cell lung cancer: inactivation of Trp53 and Rb1 in distinct cell types of adult mouse lung. *Cancer Cell* **19**: 754–764 (2011).

3.-W.D. Travis et al., WHO Classification of Tumours of the Lung, Pleura,Thymus and Heart. Fourth edition - WHO - OMS -. Fourth edition. Lyon, (2015).

4.-S.D. Castillo et al., Novel transcriptional targets of the SRY-HMG box transcription factor SOX4 link its expression to the development of small cell lung cancer. *Cancer Res.* **72**, 176–186 (2012).

5.-O.A. Romero et al., MAX inactivation in small cell lung cancer disrupts MYC-SWI/SNF programs and is synthetic lethal with BRG1. *Cancer Discov.* **4,** 292–303 (2014).

6.-C.M. Rudin et al., Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. *Nat. Rev. Cancer* **19**, 289-297 (2019).

7.-R. Blanco et al., A gene-alteration profile of human lung cancer cell lines. *Hum. Mut.* **30,** 1199–1206 (2009).

8.-M. Peifer et al., Integrative genome analyses identify key somatic driver mutations of small cell lung cancer. *Nat. Genet.* **44**, 1104-1110 (2012).

9.-J. George et al., Comprehensive genomic profiles of small cell lung cancer. *Nature* **524**, 47-53 (2015).

10.-A. Augert et al., MAX functions as a tumor suppressor and rewires metabolism in small cell lung cancer. *Cancer Cell.* **38**, 97-114 (2020).

11.-I. Comino-Méndez et al., Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. *Nat. Genet*. **4.** 663-667 (2011).

12.-D. Diolaiti, L. McFerrin, P.A. Carroll, R.N, Eisenman. Functional interactions among members of the MAX and MLX transcriptional network during oncogenesis. *Biochim. Biophys. Acta* **1849,** 484-500 (2015).

13.-L. Jiang et al. , Exome sequencing identifies somatic mutations of DDX3X in natural killer/T-cell lymphoma. *Nat. Genet.* **47**, 1061-1066 (2015).

14.-Cancer Genome Atlas Research Network, Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543-550 (2014).

15.-P.J. Hurlin, E. Steingrìmsson, N.G. Copeland, N.A. Jenkins, R.N. Eisenman, Mga, a dual-specificity transcription factor that interacts with Max and contains a T-domain DNA-binding motif. *EMBO J.* **18,** 7019-7028 (1999).

16.-J. Qin et al., The polycomb group protein L3mbtl2 assembles an atypical PRC1-family complex that is essential in pluripotent stem cells and early development. *Cell Stem Cell* **11,** 319-332 (2012).

17.-S. Hauri et al., A high-density map for navigating the human polycomb complexome. *Cell Rep*.**17**, 583-559 (2016).

18.-C.M. Rudin et al., Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. *Nat. Genet*. **44**, 1111-1116 (2012).

19.- P.A. Carroll, B.W. Freie, H. Mathsyaraja, R.N, Eisenman. The MYC transcription factor network: balancing metabolism, proliferation and oncogenesis. *Front Med.* **12**, 412-425 (2018).

20.-L. Kretzner, E.M. Blackwood, R.N, Eisenman. Transcriptional activities of the Myc and Max proteins in mammalian cells. *Curr Top Microbiol Immunol* **182,** 435-43 (1992).

21.-T.D. Littlewood, B.Amati, H. Land, G.I. Evan, Max and c-Myc/Max DNA-binding activities in cell extracts. *Oncogene* **7**, 1783-1792 (1992).

22.-H. Mathsyaraja et al., Max deletion destabilizes MYC protein and abrogates Eμ-Myc lymphomagenesis. *Genes Dev.* **3,** 1252-1264 (2019).

23.-B. Amati, S. Dalton, M.W. Brooks, T.D. Littlewood, G.I. Evan, H. Land, Transcriptional activation by the human c-Myc oncoprotein in yeast requires interaction with Max. *Nature* **359**, 423-426 (1992).

24.-D.Y. Mao et al., Analysis of Myc bound loci identified by CpG island arrays shows that Max is essential for Myc-dependent repression. *Curr. Biol.* **13**, 882-886 (2003).

25.-A. Sabo et al., Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis. *Nature* **511,** 488-492 (2014).

26.-S.J. Berberich, M.D. Cole, Casein kinase II inhibits the DNA-binding activity of Max homodimers but not Myc/Max heterodimers. *Genes Dev*. **6**, 166-176 (1992).

27.-H. Ogawa, K. Ishiguro, S. Gaubatz, D.M. Livingston, Y. Nakatani, A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science* **296,** 1132-1136 (2002).

28.-P. Trojer et al., L3MBTL2 protein acts in concert with PcG protein-mediated monoubiquitination of H2A to establish a repressive chromatin structure. *Mol. Cell.* **42,** 438-450 (2011).

29.-M. Endoh et al., PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes. *Elife* **6** (2017).

30.-B. Stielow, F. Finkernagel, T. Stiewe, A. Nist, G. Suske, MGA, L3MBTL2 and E2F6 determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6. *PLoS Genet*. **14,** e1007193 (2018).

31.-M.D. Borromeo et al., ASCL1 and NEUROD1 reveal heterogeneity in pulmonary neuroendocrine tumors and regulate distinct genetic programs. *Cell Rep.* **16**, 1259-1272 (2016).

32.-G. Mollaoglu et al., MYC drives progression of Small Cell Lung Cancer to a variant neuroendocrine subtype with vulnerability to aurora kinase inhibition. *Cancer Cell*. **31,** 270-285 (2017).

33.-P. Llabata et al., MGA suppresses the MYC pathway in lung adenocarcinoma. *Mol Cancer Res.* **18,** 574-584 (2020).

34.-A. Suzuki et al., Loss of MAX results in meiotic entry in mouse embryonic and germline stem cells. *Nat. Commun.* **7,** 11056 (2016).

35.-S.D. Castillo et al. Gene amplification of the transcription factor DP1 and CTNND1 in human lung cancer. *J. Pathol.* **222**, 89-98 (2010).

36.-M.E. Ritchie et al., (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. **43,** e47.

37.-B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

38.- Y. Zhang et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

39.- L. Zhu et al., ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11,** 237 (2010).

40.-B. Gel, A. Diez-Villanueva, E. Serra, M. Buschbeck, M.A. Peinado, R. Malinverni, regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* **32,** 289-291 (2016).

41.-F. Ramírez et al., deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res***. 44,** w160-165 (2016).

42.-S. Heinz et al., Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell.* **38**, 576–589 (2010).

ACKNOWLEDGEMENTS

We thank Isabel Bartolessis for technical assistance, the confocal facility (Carmen Casals) and the RNA-sequencing genomic facility (Anna Esteve) for important services. **Funding:** This work was supported by Spanish grant SAF2017-82186-RAEI/FEDER (UE) (to MSC) from the MINECO and a grant from the Fundación Científica Asociación Española Contra el Cancer-GCB14142170MONT (to MSC). M Torres-Diz was supported by the FPI-fellowship: BES-2012-054579, P Llabata by the FPI-fellowship: BES-2015-072204 and by the European Association for Cancer Research Travel fellowship and L Tomas-Daza by the FPI-fellowship (number PRE2019-088005). BMJ is funded by Spanish Ministry of Science, Innovation, and Universities (MICINN) project number RTI2018-094788-A-I00 and by La Caixa Banking Foundation Junior Leader project (LCF/BQ/PI19/11690001). The proteomics analyses were performed in the IJC Proteomics Unit. The IJC Proteomics Unit is part of the Spanish Platform of Molecular and Bioinformatics Resources (ProteoRed), Instituto de Salud Carlos III (PT13/0001). **Competing Interests:** All authors declare no competing financial interests.

Figure Legends

Fig. 1. SCLC cell lines with MAX inactivation have ASCL1 or NEUROD1 or combined ASCL1/NEUROD1 characteristics. (A) Western blot showing basal levels of the MYC family of proteins and MAX proteins in the different MAX-deficient SCLC cell lines, as indicated. ACTIN, protein-loading control. SCLC cell lines with amplification at *MYC* (H82), *MYCN* (H69), and *MYCL1* (HCC33 and H1963) are also included for comparison. The *MYCL1* gene is fused with the *RFL* gene in the H1963 cells, rendering a larger protein. (B) Heatmap using ASCL1 and NEUROD1 gene signatures (from reference 31) in the indicated SCLC cell lines. The gene expression has been gathered from RNA-sequencing (Lu134 and Lu165 from current work and the rest from the Cancer Cell Line Encyclopedia-CCLE). Dendrogram on the top reflects clustering of the SCLC cell lines. ASCL1-high (red shading), NEUROD1-high (green shading), POU2F3-high (gray shading) and YAP1-high (dark gray shading) groups are indicated. The expression levels of NEUROD1, ASCL1, POU2F3 and YAP1 is indicated below the dendrogram. The genetic status (*MAX, MYCN, MYC, MYCL1, MGA* and *SMARCA4*) of each cell line is also indicated with a color code. On the right, position of the NEUROD1 (green), ASCL1 (red) or both (orange) transcription factors targets in the heatmap. (C) Western blot showing basal levels of the ASCL1, NEUROD1, YAP1 and POU2F3 factors in the indicated SCLC cell lines. ACTIN, protein-loading control. (D) mRNA levels, from the RNA-seq analysis (CCLE) of ASCL1, NEUROD1 and of selected targets, grouped by three categories, MAX-deficient, MYCN/MYCL1- and MYC-amplified cells. Lines show mean; Values from each cell line are represented. *P<0.05, **P<0.01, ****P <0.001, n.s. not significant; two-sided unpaired student's t-test (*MAX*-mutant versus *MYC*-amplified groups).

Fig. 2. ChIP-seq of MYC and MAX in the distinct MAX-deficient cells and genetic backgrounds. (A) Western blot of total lysates to show the levels of MYC and MAX proteins in the indicated cells carrying ectopic overexpression of MYC ((hi)MYC) and of MYC and MAX simultaneously ((hi)MYC/MAX), at different concentrations of doxycycline (Dox). (B) Left

694 panels: heatmaps representing the normalized ChIP-seq intensities of ectopic MYC in Lu134-
695 (hi)MYC cells (1000 ng/ml Dox) and in the (hi)MYC/MAX cells (1000 ng/ml Dox). Right
696 panels: read count frequency of the heatmaps, at ±3 Kb regions centered over the TSS, of MYC
697 occupancy in the indicated cell models. (C) Representative snapshots from IGV, of ChIP-seq
698 profiles at selected target loci, performed in the indicated cell models. (D) Western blot showing
699 the ectopic expression of MAX ((lo)MAX cells (10 and 5 ng/ml Dox, Lu134 and Lu165,
700 respectively); (hi)MAX cells (1000 ng/ml Dox) and the levels of the MYC family of proteins in
701 the indicated cells. ACTIN protein-loading control. The H82 and the H69 cell lines are included
702 as a control of a *MYC*- and *MYCN*-overexpressing cells, respectively. (E) Genome-wide
703 functional annotations for peaks generated by the ChIP-seq analyses. Promoters are defined as
704 the regions ± 3Kb around the annotated TSS. (F) Venn diagrams representing the overlap of
705 MAX peaks in the Lu134 and Lu165 cells following expression of high ((hi)MAX) or
706 endogenous-like ((lo)MAX) levels of MAX.

707

708 Fig. 3. Changes in gene expression upon MAX restitution are inversely correlated with changes
709 in (hi)MYC/MAX-expressing cells. (A) Heatmap and dendrograms, using the 500 most
710 dynamic genes that changed expression in the RNA-seq, reflecting the gene expression profiles
711 of the indicated cell lines. (B) Graphs showing gene expression values in transcripts per million
712 (TPM) for the 100 most upregulated and downregulated genes selected from the (hi)MAX
713 expressing cells (from ***Dataset S1***) for each SCLC cell line and genetic context (Mock,
714 (hi)MAX, (lo)MAX or (hi)MYC/MAX). Bars show mean ± s.e.m; Two-sided unpaired
715 student's t-test, *P < 0.05; ** P < 0.01; ***P < 0.005; ****P< 0.001. (C) The common gene
716 ontology (GO) categories, of the 20 GO most enriched categories, for the upregulated (orange)
717 and downregulated (blue) genes in each of the (hi)MAX (bars on the left) and (hi)MYC/MAX
718 cells (bars on the right) (genes from ***Dataset S1*** and ***SI Appendix Fig. S5***). (D) Gene
719 enrichment set analysis (GSEA) comparing our mRNA-seq data (query datasets) with datasets
720 GSE6077 and GSE10954, from lungs of mice overexpressing nMyc and cMyc, respectively.

721 Panel bellow show in detail two selected comparatives. (E) Volcano plots depicting n-fold
722 change in gene expression for the genes bound by either MAX or MYC, in each cell type.
723 Colored dots represent the genes upregulated (in yellow) and downregulated (in blue) in each
724 cell line, among the promoters that recruit MAX or MYC (from ***Dataset S1*** and ***SI Appendix***
725 ***Fig. S5***). Changes in gene expression among genes that do not recruit MAX or MYC are
726 indicated in grey. The percentage of bound promoters among the upregulated (in yellow) and
727 downregulated (in blue) genes is also indicated. *P*- values were determined by Pearson's chi-
728 square test. *P < 0.05; ** P < 0.01; ***P < 0.005; **** P< 0.001.

729

730 Fig. 4. ncPRC1.6 can be formed in the absence of MAX. (A) Schematic representation of the
731 main components of ncPRC1.6. (B) Left panels, density plots, from the IP-MS results, showing
732 the proteins that form stable complexes with MGA in the Lu134 and Lu165 cells without MAX
733 (Mock) or in cells that express ectopic and endogenous-like levels of MAX ((lo)MAX) (see also
734 ***Dataset S3*** for detailed information). The proteins from ncPRC1.6 are indicated. The IP-MS
735 results were analyzed based on the enrichments of peptides precipitated by anti-MGA antibodies
736 relative to IgG controls (transforming 0 to 0.1), for each indicated cell line and condition. (C)
737 Confirmation of the MGA-containing complexes in each cell line and condition. MGA was
738 immunoprecipitated from whole extracts, followed by immunoblot of the indicated proteins.
739 HDAC4 was included as a negative control.

740

741 Fig. 5. MAX restitution enhances the recruitment of MGA to the DNA and represses cell
742 division- and germ cell-related functions. (A) Genome-wide functional annotations for peaks
743 generated from the ChIP-seq analyses. Promoters are defined as the regions ± 3 Kb around the
744 annotated TSS. (B) Percentage overlap of peaks at promoter regions of ChIP-seq proteins and
745 cell line models. (C) Read count frequency of the binding of MGA, among the genes
746 upregulated or downregulated in each condition (from ***Dataset S1***) ± 3 Kb regions centered over
747 the TSS, of the MGA occupancy, in each indicated cell model. (D) Venn diagrams representing

748 the overlap of MGA peaks in the Lu134 and Lu165 cells following expression of (hi)MAX or
749 (lo)MAX. The white and grey areas represent the BS2- and BS1-associated promoters,
750 respectively. (E) Violin plots representing the changes in gene expression (TPM, transcripts per
751 million), relative to the Mock cells, in each cell model and group of MGA-bound promoters (top
752 10% each of BS1 and BS2). Some of the upregulated or downregulated transcripts are indicated.
753 (F) Left panels, heatmaps representing the normalized ChIP-seq intensities for the MAX, MGA
754 and MYC proteins, in the BS1 and BS2, ranked by the intensity of the MGA binding, centered ±
755 3 Kb around the TSS. On the right, the colored bars indicate the ChIP-seq (MGA, in blue;
756 MYC, in red) with greater intensity of binding in each of the regions. Middle panels, heatmaps
757 of the gene expression from the BS1 and BS2 (10% greater intensity) in the indicated cell lines.
758 Different regions have been labeled (groups I, II, and III) according to their profile of gene
759 expression in (hi)MYC/MAX cells, compared with (lo)MAX and (hi)MAX cells. Right panels,
760 representative integrative genomics viewer (IGV) screenshots for peaks generated by the ChIP-
761 seq analyses in each cell model (screenshots Lu134 and Lu165, left and right, respectively). The
762 group and the gene ontology (GO) analyses showing selected functions for each group are also
763 indicated.

764 Fig. 6. The generation of knockouts for MGA de-repressed transcripts related with division- and
765 germ cells**.** A, Western blot of the immunoprecipitated MGA protein and of TUBULIN, from
766 the input, as protein-loading control, in the indicated lung cancer cell lines showing the
767 knockout of MGA using 3 different single guide RNA (sgRNA). B, Real-time quantitative PCR
768 of the indicated transcripts, relative to *IPO8,* and to each corresponding non-target control, in
769 the indicated lung cancer cells infected with the sgMGA (sgMGA#3, sgMGA#4, sgMGA#6).
770 Lines show mean ± s.e.m; Values represent triplicates for each of the three different sgMGA
771 (n=9). *P<0.05, **P<0.01, ***P < 0.005; ****P <0.001; two-sided unpaired student's t-test. C,
772 Representative integrative genomics viewer (IGV) screenshots for peaks generated from the
773 ChIP-seq of E2F6 (GEO accession number: GSM1010766), MAX (MAX ChIP-seq (GEO

accession number: GSM935298) MGA and MYC (GEO accession number: GSE112188) in the A549 cells

Fig. 7. MAX restitution shifts MGA DNA-binding profile from E2F sites to E-boxes and MYC-oncogenic activation decreases the levels of *ASCL1*. (A) Enrichment, given as the abundance relative to background, of the indicated DNA motifs (E2F sites and E-boxes) in the promoters bound by MYC or MAX, of the indicated cells and conditions (HOMER). (B) Enrichment, ranked by *P*-value (P<0.01), of E2F motifs and E-boxes found in the promoters bound by MGA, in the indicated cells and conditions (HOMER). (C) Number of E2F motifs and E-boxes in BS1 and BS2 (among the 10% selection) in the indicated cells. *P*-values were determined by Pearson's chi-square test. (D,E) Representative integrative genomics viewer (IGV) screenshots for peaks generated from the ChIP-seq analyses for each cell type and set of conditions. (F) Heatmaps of the gene expression of the ASCL1-only and NEUROD1-only targets selected from reference 30 (n = 540 for ASCL1 and n = 374 for NEUROD1) among the genes upregulated or downregulated in (hi)MYC/MAX cells (from ***Dataset S1***). Selected upregulated genes from each group are indicated on the right. Those that are common for both cell models are highlighted in bold. (G) Diagram showing scenarios in which the competition for available MAX is important in cell physiological processes and cancer development.