

Epigenetic co-activation of genes *MAGEA6* and *CT-GABRA3* defines orientation of a segmental duplication on the human X chromosome

Jean S. Fain^a, Aurélie Van Tongelen^a, Axelle Lorient^a and Charles De Smet^{a,b}

^a Group of Genetics and Epigenetics, de Duve Institute, Université catholique de Louvain (UCLouvain), Brussels, Belgium

^b Corresponding author: charles.desmet@uclouvain.be

Keywords: Cancer-germline genes, *MAGEA3*, *MAGEA6*, segmental duplication, bidirectional promoter, genome misassembly

Abstract

The human genome harbors many duplicated segments, which sometimes show very high sequence identity. This may complicate assignment during genome assembly. One such example is on Xq28, where the arrangement of two recently duplicated segments varies between genome assembly versions. The duplicated segments comprise highly similar genes, including *MAGEA3* and *MAGEA6*, which display specific expression in testicular germline cells, and also become aberrantly activated in a variety of tumors. Recently, a new gene was identified, *CT-GABRA3*, the transcription of which initiates inside the segmental duplication but extends far outside. According to the latest genome annotation, *CT-GABRA3* starts near *MAGEA3*, with which it shares a bidirectional promoter. In an earlier annotation however, the duplicated segment was positioned in the opposite orientation, and *CT-GABRA3* was instead coupled with *MAGEA6*. To resolve this discrepancy, and based on the contention that genes connected by a bidirectional promoter are almost always co-expressed, we decided to compare the expression profiles of *CT-GABRA3*, *MAGEA3*, and *MAGEA6*. We found that in tumor tissues and cell lines of different origins, the expression of *CT-GABRA3* was better correlated with that of *MAGEA6*. Moreover, in a cellular model of experimental induction with a DNA demethylation agent, activation *CT-GABRA3* was associated with that of *MAGEA6*, but not with that of *MAGEA3*. Together these results support a connection between *CT-GABRA3* and *MAGEA6*, and illustrate how promoter-sharing genes can be exploited to resolve genome assembly uncertainties.

Introduction

Analysis of the human genome has revealed a large number of segmental duplications typically ranging from 1-200Kb, which are either dispersed or arranged in tandem (Bailey et al. 2001). Several of these duplications occurred recently during evolution, generating segments with high (up to 99%) sequence identity (Bailey et al. 2002). Due to such high sequence similarities, recently duplicated DNA segments may be difficult to arrange precisely during genome assembly (Cheung et al. 2003).

A recent segmental duplication has been described in Xq28, comprising two near-identical segments of ~60Kb (Fig. 1). The segments are arranged in tandem and are oriented in opposite directions. They both contain a number of highly similar genes, including *MAGEA3* and *MAGEA6*, which qualify as “cancer-germline” (CG) genes. CG genes belong to a particular group of genes that normally show specific expression in testicular germline cells, but often become aberrantly activated in a wide variety of tumors (Van Der Bruggen et al. 2002). Importantly, a recent study showed that CG genes in the Xq28 duplicated segments are of clinical significance, as their activation in tumors predicts resistance to anti-CTL-A4 immunotherapy of cancer (Shukla et al. 2018). Tumoral activation of CG genes has been ascribed to a process of DNA demethylation (De Smet and Lorient 2013). It has been shown indeed that these genes rely primarily on promoter CpG methylation for repression in somatic tissues (Cannuyer et al. 2013). CG genes are therefore often co-activated in tumors that have undergone a process of global genome demethylation (Kosłowski et al. 2004).

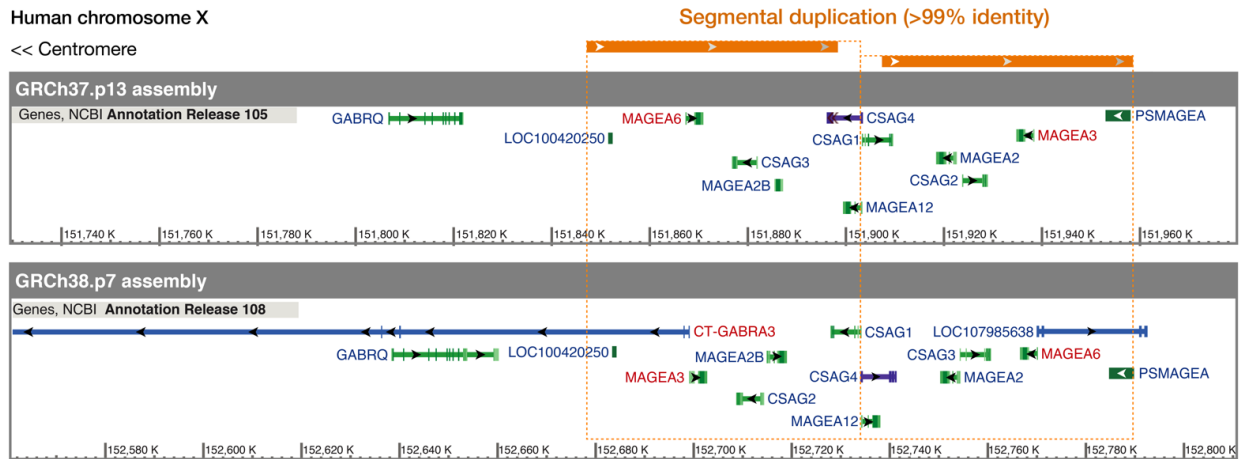


Figure 1. Uncertain assembly of a segmental duplication in Xq28. A map of the genomic Xq28 region containing genes *MAGEA3*, *MAGEA6* and *CT-GABRA3*, was generated through the Map Viewer of NCBI. Two assembly versions are depicted (GRCh37.p13, and GRCh38.p7), with positions of segmental duplications (provided by the Eichler lab). The recently characterized *CT-GABRA3* transcript, which appeared only in the latest genome annotation release, extends towards the centromere up into the *GABRA3* gene (not included in the genomic portion depicted).

In the latest reference genome (GRCh38), the segment comprising *MAGEA3* was oriented toward the centromere. This, however, was not the case in the previous reference genome version (GRCh37), where the segment oriented toward the centromere was the one containing *MAGEA6* (Fig. 1). These contradictory annotations illustrate the difficulty to correctly position near-identical segmental duplications.

Recently, we discovered a new CG gene, which we termed *CT-GABRA3* (Loriot et al. 2014). *CT-GABRA3* is a transcript variant of the brain-specific gene *GABRA3* (Gamma-Aminobutyric Acid Type A Receptor $\alpha 3$ Subunit), located in Xq28. The *CT-GABRA3* variant originates from an alternative promoter located ~250 kb upstream of the canonical *GABRA3* transcription start site. *CT-GABRA3* displays typical features of a CG gene, as it is expressed in testis, and becomes aberrantly activated in tumors such as melanoma and lung cancer (Loriot et al. 2014). According to the latest genome assembly GRCh38, *CT-GABRA3* initiates very close to *MAGEA3*, but is transcribed in the opposite direction over a long distance (~550 kb) that extends outside the segmental duplication (Fig. 1). *CT-GABRA3* and *MAGEA3* are separated by less than one hundred base pairs, and are under the influence of the same bidirectional promoter. When considering the earlier GRCh37 assembly however, *CT-GABRA3* would instead be coupled with *MAGEA6*, as the entire segmental duplication is in the reverse orientation (Fig. 1). To clarify this genome annotation uncertainty, we decided to compare the expression of *CT-GABRA3*, *MAGEA3* and *MAGEA6* in a series of melanoma samples and lung cancer cell lines, as well as in a cellular model where these genes were experimentally activated upon exposure to the demethylating agent 5-Aza-2'-deoxycytidine (5-azadC). The idea was that *CT-GABRA3* would exhibit better correlation with the gene (either *MAGEA3* or *MAGEA6*) with which it is actually coupled via a bidirectional promoter. Genes with a bidirectional promoter are indeed often co-expressed (Trinklein et al. 2004).

Material and methods

RNA-seq analysis of tumor samples and cell lines.

RNA-seq data of human melanoma samples provided by The Cancer Genome Atlas (TCGA, n=472) were downloaded from the OASIS-genomics platform (n=356) (Cancer Genome Atlas 2015; Fernandez-Banet et al. 2016). As there is no specific annotation for the *CT-GABRA3* transcript variant of *GABRA3* in OASIS expression analyses, we examined RNA-seq data by exon quantification, using the NCI-GDC Legacy Archive data portal (Grossman et al. 2016). This led to the selection of 343 melanoma samples for which we could confirm either lack of expression of any *GABRA3* variant or expression of the *CT-GABRA3* variant, which is characterized by lack of the canonical exon 1 of *GABRA3* (Loriot et al. 2014). Analysis of the distribution of *MAGEA3*, *MAGEA6* and *CT-GABRA3* expression levels among these samples was performed to define gene activation thresholds (see supplemental Fig. S1).

Study of *MAGEA3*, *MAGEA6*, and *CT-GABRA3* expression in melanoma cell lines (SKCM) was conducted by analyzing RNA-seq data from the Cancer Cell line Encyclopedia of the Broad Institute (CCLE, n=56) (Barretina et al. 2012). For the non-small-cell lung carcinoma (NSCLC, n=119), we exploited RNA-seq data of CCLE (n=93) and of the Database of Transcriptional Start Sites (DBTSS, n=26) (Suzuki et al. 2014). A description of all examined cell lines is provided in supplemental table S1. RNA-seq raw files of CCLE were downloaded from the Sequence Read Archive (SRA, accession number PRJNA523380), and those of DBTSS from the DNA Data Bank of Japan (DDBJ, accession number PRJDB2256). FASTQ files were mapped to the human reference genome GRCh37 using HISAT2-2.1.0 with default parameters. Non-unique mapping reads were removed with Samtools 0.1.19 using -q option. StringTie 1.3.4 with de novo mode (-G and -o option) was used to assemble and quantify full-length transcripts in each cell line. *CT-GABRA3*, *MAGEA3* and *MAGEA6* expression levels (transcripts per kilobase million, TPM) represent the sum of corresponding transcript variants.

Cell culture and 5-azadC treatment.

The LB2667-MEL cell line was derived from a cutaneous human melanoma at the Ludwig Institute for Cancer Research, Brussels branch (Brasseur, 1999). Cells were cultured in IMDM medium (Isocove's Modified Dulbecco's Media, Life Technologies) supplemented with 1x non-essential amino acids (NEAA), 10% fetal calf serum (FBS: Fetal Bovine Serum, Hyclone), and 100U/ml Penicillin and Streptomycin (Life Technologies). They were incubated at 37°C in a humidified atmosphere of 8% CO₂. For passages, cells were rinsed with PBS, and detached with trypsin for 5 minutes at 37°C. For treatment with 5-azadC, cells were cultured in the presence of 2µM 5-azadC (Sigma-Aldrich Chemie GmbH). After 4 days, cells received fresh 5-azadC-supplemented medium, and at day 6 the drug-containing medium was replaced by normal medium. Limiting dilution experiments were carried out at day 10 in 96-well cell culture plates, in which wells were seeded with 100 µl of cell suspension at a concentration of 30, 15 or 10 cells/ml. Plates were incubated and clones reaching confluency were transferred to larger culture dishes before RNA extraction.

RT-PCR analyses.

Total RNA samples were extracted using Tripure Isolation reagent (Roche Applied Science). RT-PCR reactions for the expression analysis of *MAGEA3*, *MAGEA6* and *CT-GABRA3* were performed as previously described (De Plaen et al. 1994; Lorient et al. 2014). PCR primers were: *MAGEA3*, 5'-TGGAGGACCAGAGGCCCCC (Fwd) and 5'-GGACGATTATCAGGAGGCCTGC (Rev); *MAGEA6*, 5'-TGGAGGAACAGAGGCCCCC (Fwd) and 5'-CAGGATGATTATCAGGAAGCCTGT (Rev); *CT-GABRA3* (Genbank #KJ620007), 5'-GGAGGCGGAGATTGCACA (Fwd) and 5'-CATCATGCCATGTCTGCCGAAA (Rev). PCR products were separated by electrophoresis on an agarose gel (Eurogentec), and visualized by ethidium bromide staining.

Results

Expression of CT-GABRA3 correlates better with that of MAGEA6 in tumor cells

To compare the expression of *CT-GABRA3*, *MAGEA3* and *MAGEA6* in tumors, we first analyzed public RNA-seq data provided by The Cancer Genome Atlas (TCGA) project (Cancer Genome Atlas 2015). We focused our analysis on melanoma samples, a tumor type where CG genes show high frequency of activation (Lorient et al. 2014; Van Der Bruggen et al. 2002). Although there was no specific annotation for the *CT-GABRA3* variant in available RNA-seq data, exon quantification through the NCI-GDC data portal allowed us to specify samples that express this mRNA variant of *GABRA3* (Grossman et al. 2016). As for *MAGEA3* and *MAGEA6*, differences in the nucleotide sequence in exons 2 and 3 makes it possible to distinguish between these two RNA species in RNA-seq data. Based on the thresholds we defined (supplemental Fig. S1A), activation of *MAGEA3*, *MAGEA6* and *CT-GABRA3* was detected in 56%, 53%, and 50% of the 343 melanoma samples, respectively. These frequencies are comparable to those reported in previous studies (Lorient et al. 2014; Van Der Bruggen et al. 2002). In most cases, we observed concurrent activation of *MAGEA3* and *MAGEA6* (Fig. 2A), confirming the high tendency of co-activation these CG genes in tumors. In a few samples however, we observed activation of either *MAGEA3* or *MAGEA6*. Importantly, analysis of *CT-GABRA3* expression in these samples revealed tighter co-occurrence of this gene with *MAGEA6* (9/12) rather than with *MAGEA3* (3/22) (Fisher exact probability test: $p=6.5 \times 10^{-4}$; Fig. 2A).

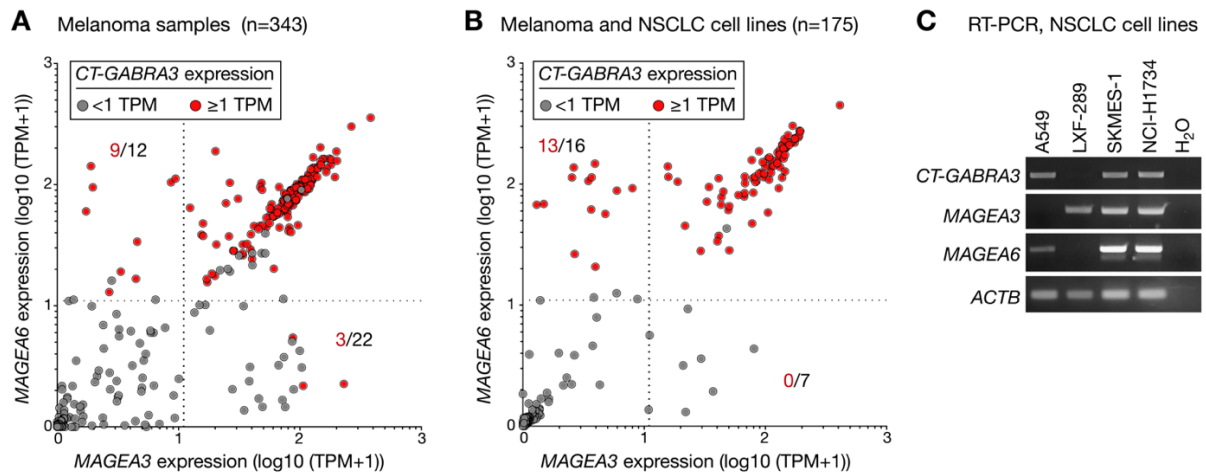


Figure 2. Expression of *CT-GABRA3* correlates better with that of *MAGEA6* in tumor cells. **A)** RNA-seq data from melanoma tissues (TCGA, n=343) were analyzed through OASIS. Samples were plotted according to *MAGEA3* and *MAGEA6* expression levels, and colored according to the presence or absence of *CT-GABRA3* activation (see supplemental Fig. S1 for definition of activation thresholds). The fraction of *CT-GABRA3*-positive samples among samples that scored positive for only *MAGEA3* or *MAGEA6* is indicated (dotted lines indicate threshold levels of activation, see supplemental Fig. S1). **B)** Similar expression analyses were performed on RNA-seq data from melanoma and non-small cell lung carcinoma (NSCLC) cell lines (CCLE and DBTSS data resources), which were processed using the StringTie *de novo* assembly algorithm. **C)** RT-PCR experiments were performed on four of the NSCLC cell lines analyzed in panel B. RNA-seq analyses predicted varying patterns of gene activation in these cell lines: A549 (*MAGEA6*+, *CT-GABRA3*+), LXF-289 (*MAGEA3*+, *CT-GABRA3*+), SKMES-1 and NCI-H1734 (*MAGEA3*+, *MAGEA6*+, *CT-GABRA3*+). These patterns were confirmed by RT-PCR. Analysis of Actin-β mRNA expression (*ACTB*) served as an internal control.

To further explore the expression profiles of *CT-GABRA3*, *MAGEA3* and *MAGEA6* in tumor cells, we resorted to the Cancer Cell line Encyclopedia of the Broad Institute (CCLE), and to the Database of transcription start sites (DBTSS), which provide RNA-seq data for tumor cell lines of various histological origins (Barretina et al. 2012; Suzuki et al. 2014). We focused our analyses on cell lines derived from melanoma (n=56), and non-small cell lung carcinoma (n=119, including adenocarcinomas and squamous cell carcinomas), where frequent activation of CG genes has been reported (Loriot et al. 2014; Van Der Bruggen et al. 2002). Access to raw RNA-seq files allowed us to define specific mRNA expression levels not only for *MAGEA3* and *MAGEA6*, but also for the *CT-GABRA3* transcript variant. The frequency of activation that we detected for these CG genes ranged between 66% and 75% in melanoma cell lines, and between 38% and 41% in non-small cell lung carcinoma cell lines (supplemental Fig. S1B-D). In both types of cell lines, we detected a high rate of co-activation of all three genes (Fig. 2B, and supplemental Fig. S2). Several cell lines however, expressed only one of the two *MAGEA* genes: either *MAGEA3* (n=7) or *MAGEA6* (n=16) (Fig. 2B, and Fig. S2). Examination of *CT-GABRA3* expression in these discriminating cell lines revealed a strong bias towards its activation in *MAGEA6*-expressing cells (13/16), rather than *MAGEA3*-expressing cells (0/7) (Fisher exact probability test, $p=4.9 \times 10^{-4}$; Fig. 2B, and Fig. S2). To confirm these *in silico* observations, RT-PCR experiments were performed on four of the non-small cell lung carcinoma cell lines that were available in the lab. According to RNA-seq data, two of these cell lines (SKMES-1 and NCI-H1734) express all three genes, one (A549) expresses *MAGEA6* and *CT-GABRA3* but not *MAGEA3*, and another (LXF-289) expresses only *MAGEA3* (supplemental Fig. S2B). RT-PCR experiments confirmed this expression pattern (Fig. 2C), thereby validating our interpretation of RNA-seq data. Together these results revealed that the expression of *CT-GABRA3* in cancer cells is highly correlated with that of *MAGEA6*, and less so with *MAGEA3*.

***CT-GABRA3* and *MAGEA6* are co-activated by a DNA methylation inhibitor**

Previous studies showed that the expression of many CG genes, including *MAGEA3*, *MAGEA6* and *CT-GABRA3*, can be induced in cells exposed to the DNA methylation inhibitor 5-azadC (Loriot et al. 2014; Sigalotti et al. 2004). It was observed that in cell populations exposed to the drug, not all CG genes become activated in every cell, and that the treatment therefore results in cellular clones with a heterogeneous pattern of CG gene activation (De Smet et al. 1999). It is expected however, that CG genes that share the same bidirectional promoter would always be co-activated upon treatment with 5-azadC.

We therefore decided to assess the extent of co-activation of *MAGEA3* or *MAGEA6* with *CT-GABRA3*, in cell clones that were derived from a 5-azadC-treated tumor cell line (LB2667-MEL), where all three genes were initially unexpressed. Thus, the cell line was treated during 6 days with 5-azadC, and was subsequently submitted to a limiting dilution to generate independent clones (Fig. 3A). RT-PCR experiments were performed on 33 clones, and revealed activation of *CT-GABRA3* in 24% of them (Fig. 3B). Analysis of the expression of *MAGEA3* and *MAGEA6* in the 33 clones showed that only the expression of *MAGEA6* was perfectly correlated with that of *CT-GABRA3* (Fig. 3B). *MAGEA3* instead, showed activation in 4 clones that did not express *CT-GABRA3*, and remained silent in 6 clones that expressed *CT-GABRA3*. These observations confirmed that the activation of *CT-GABRA3* is more strongly associated with that of *MAGEA6*, as compared with that of *MAGEA3* (Fisher exact probability test, 4.8×10^{-3} ; Fig 3C).

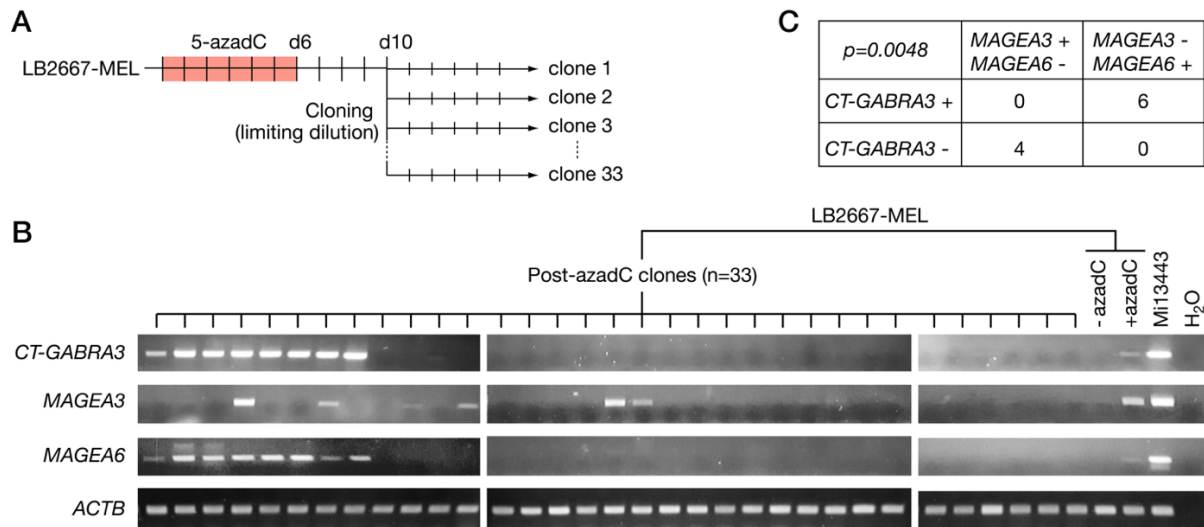


Figure 3. Epigenetic co-activation of *CT-GABRA3* and *MAGEA6*. **A**) Schematic outline of the experiment of 5-azadC activation. LB2667-MEL cells (initially negative for *CT-GABRA3*, *MAGEA3* and *MAGEA6* expression) were treated with 2 μ M 5-azadC during 6 days. At day 10, post-azadC clones were derived by limiting dilution. **B**) Activation of *CT-GABRA3*, *MAGEA3* and *MAGEA6* expression in post-azadC clones was assessed by RT-PCR. Expression was also tested in the total LB2667-MEL cell population after 6 days of exposure to 5-azadC or control vehicle. Mi13443 melanoma cell line, which expresses all three genes constitutively, was used as a positive control. *ACTB* was amplified to confirm equal cDNA amounts in all samples. **C**) Expression patterns, which we observed in post-azadC clones, were reported in a 2x2 contingency table, and a Fisher's exact test was applied.

Discussion

Annotation uncertainties persist in the human genome, in particular within regions that contain highly similar duplicated segments. This is the case in human Xq28, where the orientation of two nearly identical segments varies according to the genome assembly version. The two segments comprise highly similar genes, including *MAGEA3* and *MAGEA6*. Recently, we identified a new gene, *CT-GABRA3*, which initiates transcription from a bidirectional promoter located within the segmental duplication and extends outside towards the centromere. We decided to take advantage of this particular configuration to clarify the exact orientation of the duplicated segments.

Expression of *CT-GABRA3*, *MAGEA3*, and *MAGEA6* was evaluated in a large set of tumor tissues and cell lines, as well as in a cellular model of experimental induction. Of note, activation frequencies and correlations of the three X-linked CG genes in cancer tissues and cell lines were not affected by gender (supplemental Fig. S3). Together our data indicate that the activation of *CT-GABRA3* is better correlated with that of *MAGEA6* than with that of *MAGEA3*. We conclude therefore that *MAGEA6*, but not *MAGEA3*, is linked with *CT-GABRA3* through a bidirectional promoter. This is in favor of the former GRCh37 assembly, where the duplicated segment containing *MAGEA6* in Xq28 was oriented towards the centromere (Fig. 1).

The involvement of a process of global genome demethylation in the activation of CG genes explains their frequent co-expression in tumors (De Smet and Lorient 2013). *MAGEA3* and *MAGEA6/CT-GABRA3* genes, in

particular, show a higher propensity of co-activation, as compared with other CG genes. This may be explained by their proximity on the genome, and the high sequence identity within their 5'-region (supplemental Fig. S4), which stems from a recent gene conversion event (Katsura and Satta 2011). The occurrence of several tumors that express only *MAGEA3* and not *MAGEA6/CT-GABRA3*, or *vice versa*, may possibly be attributed to the stochastic nature of epigenetic alterations (Feinberg 2014).

More generally, we believe that our work illustrates how the presence of a bidirectional promoter can be exploited to define the position of a DNA segment. Considering the high number of bidirectional promoters in genomes, which initiate transcription of either messenger or non-coding RNAs (Wei et al. 2011), it is conceivable that the procedure we described in the present study can be applied to other loci, and will help to resolve a number of genomic annotation uncertainties.

Acknowledgments

This work was supported by grants from the D.G. Higher Education and Scientific Research of the French Community of Belgium (Action de Recherches Concertées) and from the Fonds special de recherche (FSR) of the UCLouvain, Belgium. AVT was a recipient of a Télévie grant from the FRS-FNRS, Belgium [#7.4581.13]. AL is supported by the de Duve Institute, Brussels, Belgium.

References

- Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE: Recent segmental duplications in the human genome. *Science* 297:1003-1007 (2002).
- Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE: Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res* 11:1005-1017 (2001).
- Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, Berger MF, Monahan JE, Morais P, Meltzer J, Korejwa A, Jane-Valbuena J, Mapa FA, Thibault J, Bric-Furlong E, Raman P, Shipway A, Engels IH, Cheng J, Yu GK, Yu J, Aspesi P, Jr., de Silva M, Jagtap K, Jones MD, Wang L, Hatton C, Palessandolo E, Gupta S, Mahan S, Sougnez C, Onofrio RC, Liefeld T, MacConaill L, Winckler W, Reich M, Li N, Mesirov JP, Gabriel SB, Getz G, Ardlie K, Chan V, Myer VE, Weber BL, Porter J, Warmuth M, Finan P, Harris JL, Meyerson M, Golub TR, Morrissey MP, Sellers WR, Schlegel R, Garraway LA: The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483:603-607 (2012).
- Cancer Genome Atlas N: Genomic Classification of Cutaneous Melanoma. *Cell* 161:1681-1696 (2015).
- Cannuyer J, Lorient A, Parvizi GK, De Smet C: Epigenetic hierarchy within the *MAGEA1* cancer-germline gene: promoter DNA methylation dictates local histone modification. *PLoS ONE* 8:e58743 (2013).
- Cheung J, Estivill X, Khaja R, MacDonald JR, Lau K, Tsui LC, Scherer SW: Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. *Genome biology* 4:R25 (2003).
- De Plaen E, Arden K, Traversari C, Gaforio JJ, Szikora J-P, De Smet C, Brasseur F, van der Bruggen P, Lethé B, Lurquin C, Brasseur R, Chomez P, De Backer O, Cavenée W, Boon T: Structure, chromosomal localization and expression of twelve genes of the *MAGE* family. *Immunogenetics* 40:360-369 (1994).
- De Smet C, Lorient A: DNA hypomethylation and activation of germline-specific genes in cancer. *Adv Exp Med Biol* 754:149-166 (2013).
- De Smet C, Lurquin C, Lethé B, Martelange V, Boon T: DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. *Mol Cell Biol* 19:7327-7335 (1999).
- Feinberg AP: Epigenetic stochasticity, nuclear structure and cancer: the implications for medicine. *J Intern Med* 276:5-11 (2014).
- Fernandez-Banet J, Esposito A, Coffin S, Horvath IB, Estrella H, Schefzick S, Deng S, Wang K, K AC, Ding Y, Roberts P, Rejto PA, Kan Z: OASIS: web-based platform for exploring cancer multi-omics data. *Nat Methods* 13:9-10 (2016).
- Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, Staudt LM: Toward a Shared Vision for Cancer Genomic Data. *N Engl J Med* 375:1109-1112 (2016).
- Katsura Y, Satta Y: Evolutionary history of the cancer immunity antigen *MAGE* gene family. *PLoS One* 6:e20365 (2011).
- Koslowski M, Bell C, Seitz G, Lehr HA, Roemer K, Muntefering H, Huber C, Sahin U, Tureci O: Frequent nonrandom activation of germline genes in human cancer. *Cancer Res* 64:5988-5993. (2004).
- Lorient A, Van Tongelen A, Blanco J, Klaessens S, Cannuyer J, Van Baren N, Decottignies A, De Smet C: A novel cancer-germline transcript carrying pro-metastatic miR-105 and TET-targeting miR-767 induced by DNA hypomethylation in tumors. *Epigenetics* 9:1163-1171 (2014).
- Shukla SA, Bachireddy P, Schilling B, Galonska C, Zhan Q, Bango C, Langer R, Lee PC, Gusenleitner D, Keskin DB, Babadi M, Mohammad A, Gnirke A, Clement K, Cartun ZJ, Van Allen EM, Miao D, Huang Y, Snyder A, Merghoub T, Wolchok JD, Garraway LA, Meissner A, Weber JS, Hacohen N, Neuberg D, Potts PR, Murphy GF, Lian CG, Schadendorf D, Hodi FS, Wu CJ: Cancer-Germline Antigen Expression Discriminates Clinical Outcome to CTLA-4 Blockade. *Cell* 173:624-633 e628 (2018).
- Sigalotti L, Fratta E, Coral S, Tanzarella S, Danielli R, Colizzi F, Fonsatti E, Traversari C, Altomonte M, Maio M: Intratumor heterogeneity of cancer/testis antigens expression in human cutaneous melanoma is methylation-regulated and functionally reverted by 5-aza-2'-deoxycytidine. *Cancer Res* 64:9167-9171 (2004).
- Suzuki A, Makinoshima H, Wakaguri H, Esumi H, Sugano S, Kohno T, Tsuchihara K, Suzuki Y: Aberrant transcriptional regulations in cancers: genome, transcriptome and epigenome analysis of lung adenocarcinoma cell lines. *Nucleic Acids Res* 42:13557-13572 (2014).
- Trinklein ND, Aldred SF, Hartman SJ, Schroeder DI, Otillar RP, Myers RM: An abundance of bidirectional promoters in the human genome. *Genome Res* 14:62-66 (2004).
- Van Der Bruggen P, Zhang Y, Chaux P, Stroobant V, Panichelli C, Schultz ES, Chapiro J, Van Den Eynde BJ, Brasseur F, Boon T: Tumor-specific shared antigenic peptides recognized by human T cells. *Immunol Rev* 188:51-64. (2002).
- Wei W, Pelechano V, Jarvelin AI, Steinmetz LM: Functional consequences of bidirectional promoters. *Trends Genet* 27:267-276 (2011).

Melanoma cell lines (n=56)

Cell lines	Project	Tissue	Type	Subtype	Gender
A-375	CCLE	skin	SKCM	-	female
A101D	CCLE	skin	SKCM	-	male
A2058	CCLE	skin	SKCM	-	male
C32	CCLE	skin	SKCM	-	male
CJM	CCLE	skin	SKCM	-	NA
COLO 741	CCLE	skin	SKCM	-	female
COLO 792	CCLE	skin	SKCM	-	male
COLO 829	CCLE	skin	SKCM	-	male
COLO-679	CCLE	skin	SKCM	-	female
COLO-783	CCLE	skin	SKCM	-	female
COLO-900	CCLE	skin	SKCM	-	male
G-361	CCLE	skin	SKCM	-	male
HMCB	CCLE	skin	SKCM	-	female
Hs 294.T	CCLE	skin	SKCM	-	male
Hs 800.T	CCLE	skin	SKCM	-	male
Hs 688(A).T	CCLE	skin	SKCM	-	male
Hs 695T	CCLE	skin	SKCM	-	male
Hs 834.T	CCLE	skin	SKCM	-	female
Hs 839.T	CCLE	skin	SKCM	-	female
Hs 852.T	CCLE	skin	SKCM	-	male
Hs 895.T	CCLE	skin	SKCM	-	female
Hs 934.T	CCLE	skin	SKCM	-	female
Hs 936.T	CCLE	skin	SKCM	-	male
Hs 939.T	CCLE	skin	SKCM	-	female
Hs 940.T	CCLE	skin	SKCM	-	male
Hs 944.T	CCLE	skin	SKCM	-	male
HT-144	CCLE	skin	SKCM	-	male
IGR-1	CCLE	skin	SKCM	-	male
IGR-37	CCLE	skin	SKCM	-	male
IGR-39	CCLE	skin	SKCM	-	male
IPC-298	CCLE	skin	SKCM	-	female
K029AX	CCLE	skin	SKCM	-	NA
LOX IMVI	CCLE	skin	SKCM	-	male
Malme-3M	CCLE	skin	SKCM	-	male
MDA-MB-435	CCLE	skin	SKCM	-	female
MEL-HO	CCLE	skin	SKCM	-	female
MEL-JUSO	CCLE	skin	SKCM	-	female
MeWo	CCLE	skin	SKCM	-	male
RPMI-7951	CCLE	skin	SKCM	-	female
RMH-421	CCLE	skin	SKCM	-	male
SH-4	CCLE	skin	SKCM	-	female
SK-MEL-1	CCLE	skin	SKCM	-	male
SK-MEL-24	CCLE	skin	SKCM	-	male
SK-MEL-28	CCLE	skin	SKCM	-	male
SK-MEL-3	CCLE	skin	SKCM	-	female
SK-MEL-30	CCLE	skin	SKCM	-	male
SK-MEL-31	CCLE	skin	SKCM	-	female
SK-MEL-5	CCLE	skin	SKCM	-	female
UACC-257	CCLE	skin	SKCM	-	NA
UACC-62	CCLE	skin	SKCM	-	NA
WM-115	CCLE	skin	SKCM	-	female
WM-266-4	CCLE	skin	SKCM	-	female
WM-793	CCLE	skin	SKCM	-	male
WM-88	CCLE	skin	SKCM	-	male
WM-983B	CCLE	skin	SKCM	-	male
WM1799	CCLE	skin	SKCM	-	male

NSCLC cell lines (n=119)

Cell lines	Project	Tissue	Type	Subtype	Gender	Cell lines	Project	Tissue	Type	Subtype	Gender
A427	DBTSS	lung	NSCLC	LUAD	male	NCI-H1734	CCLE	lung	NSCLC	LUAD	female
A549	DBTSS	lung	NSCLC	LUAD	male	NCI-H1755	CCLE	lung	NSCLC	LUAD	female
ABC-1	DBTSS	lung	NSCLC	LUAD	male	NCI-H1781	CCLE	lung	NSCLC	LUAD	female
CAL-12T	CCLE	lung	NSCLC	-	male	NCI-H1792	CCLE	lung	NSCLC	LUAD	male
Calu-1	CCLE	lung	NSCLC	LUSC	male	NCI-H1793	CCLE	lung	NSCLC	-	female
CALU-3	CCLE	lung	NSCLC	LUAD	male	NCI-H1819	DBTSS	lung	NSCLC	LUAD	female
COR-L105	CCLE	lung	NSCLC	LUAD	male	NCI-H1838	CCLE	lung	NSCLC	-	female
DV-90	CCLE	lung	NSCLC	LUAD	male	NCI-H1869	CCLE	lung	NSCLC	LUSC	male
EBC-1	CCLE	lung	NSCLC	LUSC	male	NCI-H1944	CCLE	lung	NSCLC	-	female
EKVX	CCLE	lung	NSCLC	LUAD	male	NCI-H1975	DBTSS	lung	NSCLC	LUAD	female
EPLC-272H	CCLE	lung	NSCLC	LUSC	male	NCI-H2009	CCLE	lung	NSCLC	LUAD	female
HARA	CCLE	lung	NSCLC	LUSC	male	NCI-H2023	CCLE	lung	NSCLC	LUAD	male
HCC-1171	CCLE	lung	NSCLC	-	male	NCI-H2030	CCLE	lung	NSCLC	-	male
HCC-1195	CCLE	lung	NSCLC	LUAD	male	NCI-H2073	CCLE	lung	NSCLC	LUAD	female
HCC-15	CCLE	lung	NSCLC	LUSC	male	NCI-H2085	CCLE	lung	NSCLC	LUAD	male
HCC-1588	CCLE	lung	NSCLC	LUSC	female	NCI-H2087	CCLE	lung	NSCLC	LUAD	male
HCC-1833	CCLE	lung	NSCLC	LUAD	female	NCI-H2106	CCLE	lung	NSCLC	-	male
HCC-2108	CCLE	lung	NSCLC	LUAD	male	NCI-H2110	CCLE	lung	NSCLC	-	NA
HCC-2279	CCLE	lung	NSCLC	LUAD	female	NCI-H2122	CCLE	lung	NSCLC	LUAD	female
HCC-2814	CCLE	lung	NSCLC	LUSC	male	NCI-H2126	DBTSS	lung	NSCLC	LUAD	male
HCC-366	CCLE	lung	NSCLC	ASC	female	NCI-H2170	CCLE	lung	NSCLC	LUSC	male
HCC-44	CCLE	lung	NSCLC	LUAD	female	NCI-H2172	CCLE	lung	NSCLC	-	female
HCC-78	CCLE	lung	NSCLC	LUAD	male	NCI-H2228	DBTSS	lung	NSCLC	LUAD	female
HCC-95	CCLE	lung	NSCLC	LUSC	male	NCI-H226	CCLE	lung	NSCLC	LUSC	male
HCC2935	CCLE	lung	NSCLC	-	male	NCI-H2291	CCLE	lung	NSCLC	LUAD	male
HCC364	CCLE	lung	NSCLC	LUAD	male	NCI-H23	CCLE	lung	NSCLC	-	male
HCC4006	CCLE	lung	NSCLC	LUAD	male	NCI-H2342	CCLE	lung	NSCLC	LUAD	male
HCC827	CCLE	lung	NSCLC	LUAD	female	NCI-H2347	DBTSS	lung	NSCLC	LUAD	female
HCC827 GR5	CCLE	lung	NSCLC	-	female	NCI-H2405	CCLE	lung	NSCLC	LUAD	male
HLF-a	CCLE	lung	NSCLC	LUSC	female	NCI-H2444	CCLE	lung	NSCLC	-	male
HOP-62	CCLE	lung	NSCLC	LUAD	female	NCI-H292	CCLE	lung	NSCLC	MEC	female
HS 229.T	CCLE	lung	NSCLC	LUAD	male	NCI-H322	DBTSS	lung	NSCLC	LUAD	male
HS 618.T	CCLE	lung	NSCLC	LUAD	female	NCI-H3255	CCLE	lung	NSCLC	LUAD	female
II18	DBTSS	lung	NSCLC	LUAD	NA	NCI-H358	CCLE	lung	NSCLC	LUAD	male
KNS-62	CCLE	lung	NSCLC	LUSC	male	NCI-H441	CCLE	lung	NSCLC	LUAD	male
LC-1F	CCLE	lung	NSCLC	LUSC	male	NCI-H520	CCLE	lung	NSCLC	LUSC	male
LC2ad	DBTSS	lung	NSCLC	LUAD	female	NCI-H522	CCLE	lung	NSCLC	-	male
LK-2	CCLE	lung	NSCLC	LUSC	male	NCI-H596	CCLE	lung	NSCLC	ASC	male
LOU-NH91	CCLE	lung	NSCLC	LUSC	female	NCI-H647	CCLE	lung	NSCLC	ASC	male
LU65	CCLE	lung	NSCLC	-	male	NCI-H650	CCLE	lung	NSCLC	LUAD	male
LUDLU-1	CCLE	lung	NSCLC	LUSC	male	NCI-H838	CCLE	lung	NSCLC	-	male
LXF-289	CCLE	lung	NSCLC	LUAD	male	NCI-H854	CCLE	lung	NSCLC	LUAD	male
MOR/CPR	CCLE	lung	NSCLC	LUAD	NA	PC-14	DBTSS	lung	NSCLC	LUAD	male
NCI-H1299	DBTSS	lung	NSCLC	LUAD	male	PC-3	DBTSS	lung	NSCLC	LUAD	female
NCI-H1355	CCLE	lung	NSCLC	LUAD	male	PC-7	DBTSS	lung	NSCLC	LUAD	male
NCI-H1373	CCLE	lung	NSCLC	LUAD	male	PC-9	DBTSS	lung	NSCLC	LUAD	male
NCI-H1385	CCLE	lung	NSCLC	LUSC	female	RERF-LC-Ad1	DBTSS	lung	NSCLC	LUAD	male
NCI-H1395	CCLE	lung	NSCLC	LUAD	female	RERF-LC-Ad2	DBTSS	lung	NSCLC	LUAD	male
NCI-H1435	CCLE	lung	NSCLC	-	female	RERF-LC-AI	CCLE	lung	NSCLC	LUSC	male
NCI-H1437	DBTSS	lung	NSCLC	LUAD	male	RERF-LC-KJ	DBTSS	lung	NSCLC	LUAD	male
NCI-H1563	CCLE	lung	NSCLC	LUAD	male	RERF-LC-MS	DBTSS	lung	NSCLC	LUAD	male
NCI-H1568	CCLE	lung	NSCLC	-	female	RERF-LC-OK	DBTSS	lung	NSCLC	LUAD	female
NCI-H1573	CCLE	lung	NSCLC	LUAD	female	RERF-LC-Sq1	CCLE	lung	NSCLC	LUSC	female
NCI-H1623	CCLE	lung	NSCLC	LUAD	male	SK-LU-1	CCLE	lung	NSCLC	LUAD	female
NCI-H1648	DBTSS	lung	NSCLC	LUAD	male	SK-MES-1	CCLE	lung	NSCLC	LUSC	male
NCI-H1650	DBTSS	lung	NSCLC	LUAD	male	Sq-1	CCLE	lung	NSCLC	LUSC	male
NCI-H1651	CCLE	lung	NSCLC	LUAD	male	SW 1573	CCLE	lung	NSCLC	LUSC	female
NCI-H1666	CCLE	lung	NSCLC	LUAD	female	SW 900	CCLE	lung	NSCLC	LUSC	male
NCI-H1693	CCLE	lung	NSCLC	LUAD	female	VMRC-LCD	DBTSS	lung	NSCLC	LUAD	male
NCI-H1703	DBTSS	lung	NSCLC	LUAD	male						

Table S1. Description of human melanoma and NSCLC cell lines. Abbreviations : Broad Institute cancer cell line encyclopedia (CCLE), Database of transcriptional start sites (DBTSS), Skin cutaneous melanoma (SKCM), Non-small-cell lung carcinoma (NSCLC), Lung adenocarcinoma (LUAD), Lung squamous cell carcinoma (LUSC), Adenosquamous carcinoma of the lung (ASC), Mucoepidermoid carcinoma of the lung (MEC).

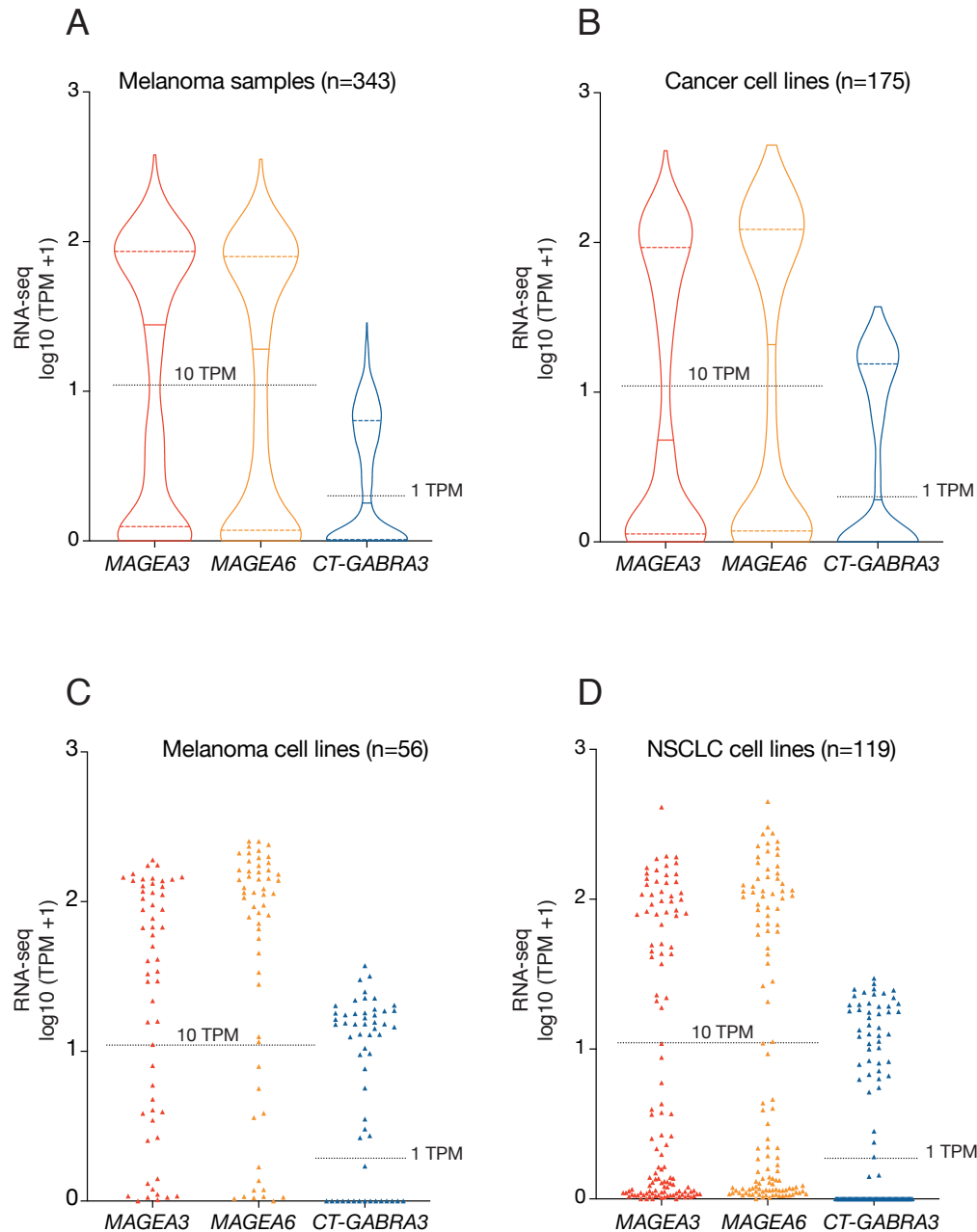
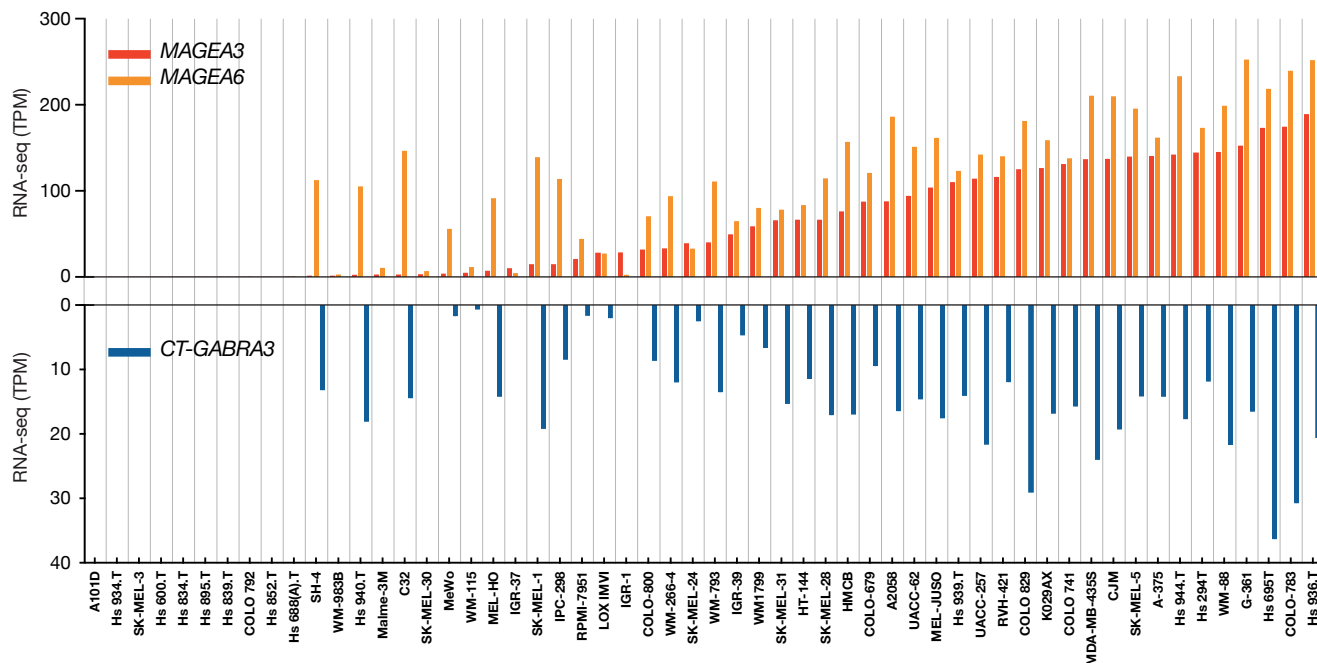


Figure S1. Distribution of *MAGEA3*, *MAGEA6* and *CT-GABRA3* mRNA expression in melanoma samples and in tumor cell lines (melanoma and NSCLC): defining gene activation thresholds. **A)** Violin plots displaying *MAGEA3*, *MAGEA6* and *CT-GABRA3* expression levels in melanoma samples (n=343, TCGA project). Expression data, which derive from RNA-seq data of the OASIS-genomics platform, are expressed as log10 (TPM + 1). Median and quartiles are indicated in the plots by continuous and discontinuous lines, respectively. For each gene, the violin plot reveals two groups of samples with respect to the expression level. Based on this we established an activation threshold for *MAGEA3* and *MAGEA6* (≥ 10 TPM), and for *CT-GABRA3* (≥ 1 TPM). **B)** Similar analyses were performed on tumor cell lines (melanoma and NSCLC), using RNA-seq raw files (n=175, CCLE and DBTSS data resource) processed with the StringTie de novo assembly algorithm. The same activation thresholds were defined. **C-D)** Dot plot representation of gene expression levels in melanoma cell lines and in NSCLC cell lines. Each dots represents a single cell line.

A Melanoma cell lines (n=56)



B NSCLC cell lines (n=119)

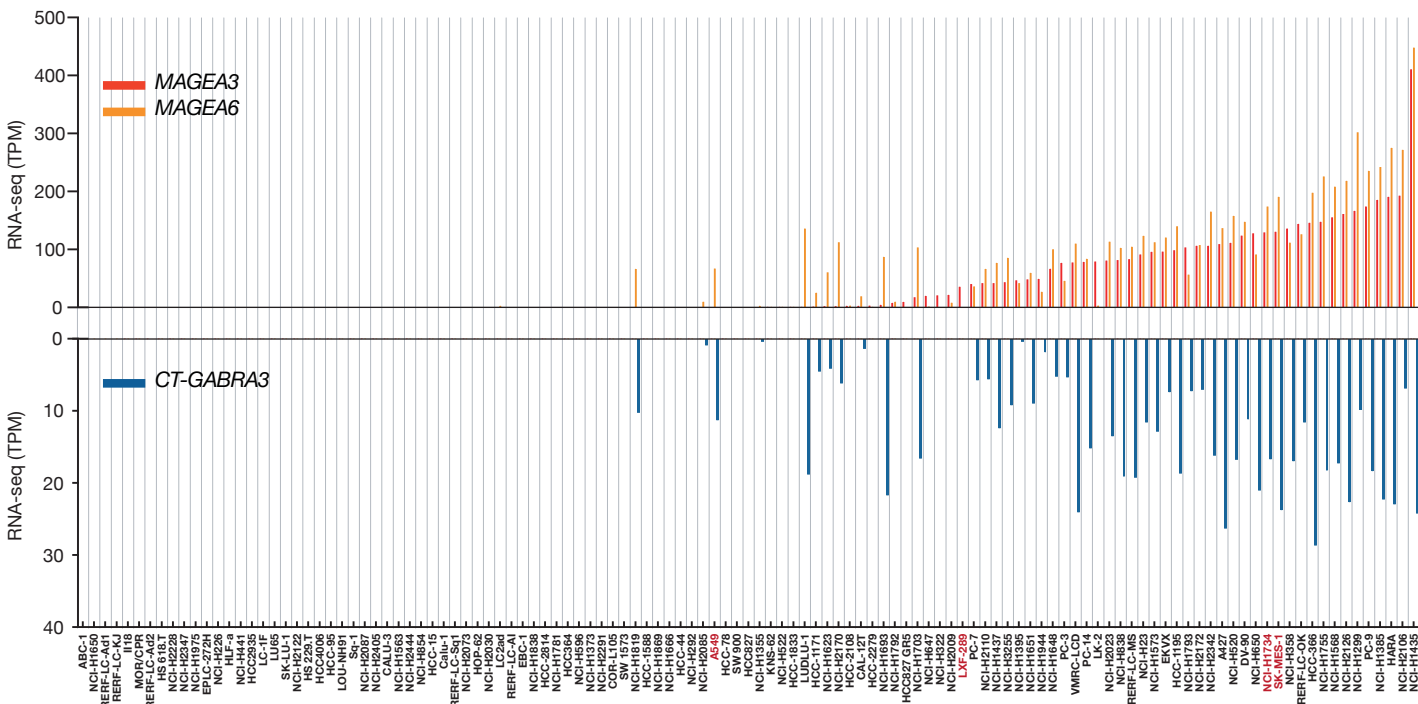


Figure S2. Expression of *MAGEA3*, *MAGEA6* and *CT-GABRA3* in melanoma and NSCLC cell lines. A) RNA-seq raw data from melanoma cell lines (n=56, CCLE data resource) were analyzed using HISAT2 alignment program and the StringTie de novo assembly algorithm. Expression levels of *MAGEA3*, *MAGEA6* and *CT-GABRA3* are depicted for each cell line (TPM). Cell lines are ordered according to *MAGEA3* expression level. **B)** Similar analyses were performed on NSCLC cell lines (n=119, CCLE and DDBJ data resource). Cell lines that were used in RT-PCR experiments (see Fig. 2C) are highlighted in red.

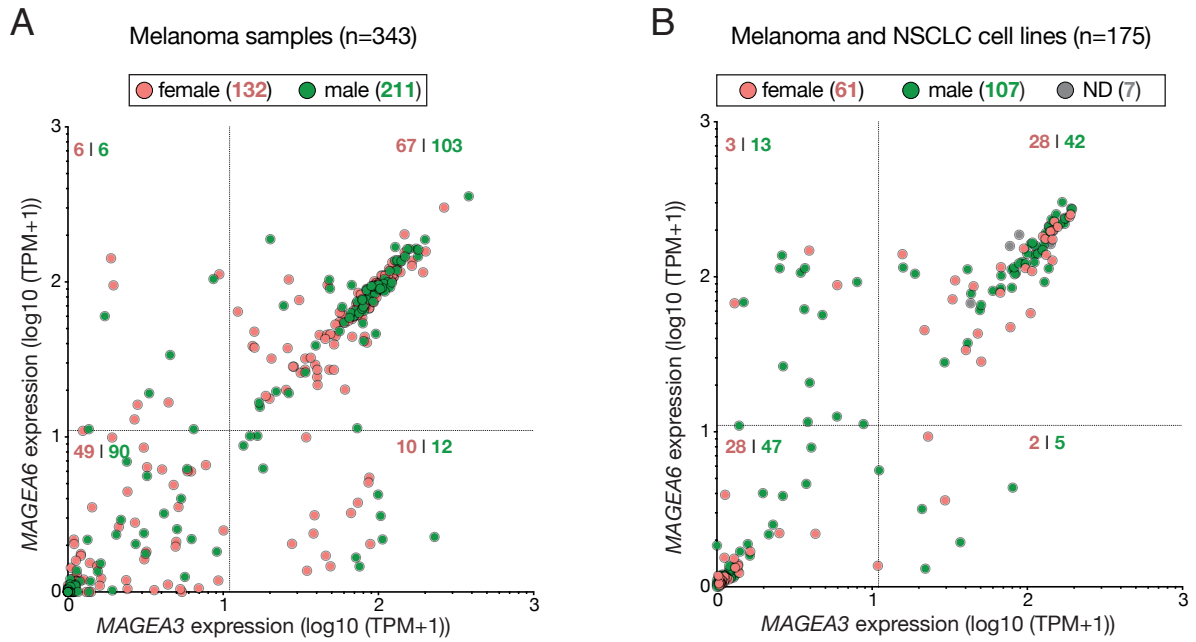


Figure S3. Activation of *MAGEA3* and *MAGEA6/CT-GABRA3* genes in tumor cells is not affected by gender. Plots depicted in Fig. 2A, and 2B are represented here with color-code identification of gender in panels A and B, respectively. No statistically significant differences was observed among the different *MAGEA3* and/or *MAGEA6* expression groups. There was also no significant difference between *CT-GABRA3*-positive and -negative samples.

- 1000 TTGTGTTTGCAGCTTGCAGTGTGGTTGGAGGGGGTTAGCAGCAGGGATGTTGGGGAGGTTTGTAGC **CG**GCCTACAC **CG**GTAGATGACAGAATGGGTAGAA
 TTGTGTTTGCAGCTTGCAGTGTGGTTGGAGGGGGTTAGCAGCAGGGATGTTGGGGAGGTTTGTAGC **CG**GCCTACAC **CG**GTAGATGACAGAATGGGTAGAA
 - 900 TAAAGTTTGAATTTTCCACTTCACCTTCTTTCACAACTCTGAGGCAGCCTCTGAAAACA **CG**ATGCCAAGAGCCCTAGGTAATAGCAGGACACCAGGAAA
 TAAAGTTTGAATTTTCCACTTCACCTTCTTTCACAACTCTGAGGCAGCCTCTGAAAACA **CG**ATGCCAAGAGCCCTAGGTAATAGCAGGACACCAGGAAA
 - 800 ATTTGGTTGGTTAAGGCAGTATGACTCCAGAGTTCCTGCTAATAACAACCTGAAACCACCATAGTGGCAGAGGAATTACATTTTAAAAAATAATCTCT
 ATTTGGTTGGTTAAGGCAGTATGACTCCAGAGTTCCTGCTAATAACAACCTGAAACCACCATAGTGGCAGAGGAATTACATTTTAAAAAATAATCTCT
 - 700 TTCAACAGAACACAAAGACTGAAATAGGAGGTCACTACAAC **CGCG**GGAGCTGC **CG**CCCG **CG**CCCTGCAGGAGACCTGGCCTGGGACC **CG**CAGGCATTCT
 TTCAACAGAACACAAAGACTGAAATAGGAGGTCACTACAAC **CGCG**GGAGCTGC **CG**CCCG **CG**CCCTGCAGGAGACCTGGCCTGGGACC **CG**CAGGCATTCT
 - 600 CTACAAGGGGTGCAGCTGTGCAATGCTCAGAGTGACAGAAACAGAGCATCTCCTGCCCATCACTTCACTCAACAGCCAGAGGTGA **CG**AAGAC **CG**ACCCT
 CTACAAGGGGTGCAGCTGTGCAATGCTCAGAGTGACAGAAACAGAGCATCTCCTGCCCATCACTTCACTCAACAGCCAGAGGTGA **CG**AAGAC **CG**ACCCT
 - 500 CCTGAGTGAGGACTGAGGGTCCACAC **CG**CCCCCCCCACCCACACACCATAGAGGGACCACAGAATCCAGCTCAGCCCTCTTGTGAGCCCTGGTAAAC **CG**C
 CCTGAGTGAGGACTGAGGGTCCACAC **CG**CCCCCCCCACCCACACACCATAGAGGGACCACAGAATCCAGCTCAGCCCTCTTGTGAGCCCTGGTAAAC **CG**C
 - 400 AGGCAGTGATGTACCCAGACCACACCCCTTCCCCCAATGCCACTTCAGGGGGACTCAGAGTCAGAGACTTGGTCTGAGGGGAGCAGAAGCAATCTGCAG
 AGGCAGTGATGTACCCAGACCACACCCCTTCCCCCAATGCCACTTCAGGGGGACTCAGAGTCAGAGACTTGGTCTGAGGGGAGCAGAAGCAATCTGCAG
 - 300 AGGATGG **CG**GTCCAGGCTCAGCCAGGCATCAACTTCAGGACCTTGAGGGATGAC **CG**AAGGCC **CG**CCCCACCCCAACTCCCC **CG**ACCCACCAGGA
 AGGATGG **CG**GTCCAGGCTCAGCCAGGCATCAACTTCAGGACCTTGAGGGATGAC **CG**AAGGCC **CG**CCCCACCCCAACTCCCC **CG**ACCCACCAGGA
 - 200 TCTACAGCCTCAGGACCC **CG**TCCCAATCCTTACCCTTGCCCATCACCATCTTCATGCTTACCTCCACCCCATC **CG**ATCCCATCCAGGCAGAATCC
 TCTACAGCCTCAGGACCC **CG**TCCCAATCCTTACCCTTGCCCATCACCATCTTCATGCTTACCTCCACCCCATC **CG**ATCCCATCCAGGCAGAATCC
 - 100 AGTTCCACCCCTGCC **CG**GAACCCAGGGTAGTAC **CG**TTGCCAGGATGTGAC **CG**CCACTGACTTGC **CG**CATTGGAGGTGAGAAGAC **CG**CGAGATTCT **CG**CCCTG
 AGTTCCACCCCTGCC **CG**GAACCCAGGGTAGTAC **CG**TTGCCAGGATGTGAC **CG**CCACTGACTTGC **CG**CATTGGAGGTGAGAAGAC **CG**CGAGATTCT **CG**CCCTG
 0 **→ MAGEA6**
 AGCAAC **CG**AGCGAC **CG**GCCTGAC **CG**TCCG **CG**GAGGGAAGC **CG**GCCCAGGCT **CG**GTGAGGAGGCAAGGTAAGA **CG**CTGAGGGAGGACTGAGG **CG**GGCCTCACCT
 AGCAAC **CG**AGCGAC **CG**GCCTGAC **CG**TCCG **CG**GAGGGAAGC **CG**GCCCAGGCT **CG**GTGAGGAGGCAAGGTAAGA **CG**CTGAGGGAGGACTGAGG **CG**GGCCTCACCT
→ MAGEA3
 + 100 CAGACAGAGGGCCTCAAATAATCCAGTGTGCTGCTGCTGC **CG**GGCCTGGGCCACCC **CG**CAGGGGAAGACTTCCAGGCTGGGT **CG**CCACTACCTCACCCC
 CAGACAGAGGGCCTCAAATAATCCAGTGTGCTGCTGCTGC **CG**GGCCTGGGCCACCC **CG**CAGGGGAAGACTTCCAGGCTGGGT **CG**CCACTACCTCACCCC
 + 200 GC **CG**ACCC **CG**CGCTTTAGCCA **CG**GGGAACCTCTGGGACAGAGCTTAATGTGGCCAGGGCAGGCTGGTTAGAAGAGGTCAGGGCCCA **CG**CTGTGGCAG
 GCG **CG**ACCC **CG**CGCTTTAGCCA **CG**GGGAACCTCTGGGACAGAGCTTAATGTGGCCAGGGCAGGCTGGTTAGAAGAGGTCAGGGCCCA **CG**CTGTGGCAG
 + 300 GAATCAAGGTCAGGACCC **CG**AGAGGGAACCTGAGGGCAGCCTAACCCACCACCTCACCACCATTC **CG**TCCCCAACACCAACCCACCCCATCCCCAT
 GAATCAAGGTCAGGACCC **CG**AGAGGGAACCTGAGGGCAGCCTAACCCACCACCTCACCACCATTC **CG**TCCCCAACACCAACCCACCCCATCCCCAT
 + 400 TCCCCATTCCTATCCCCACCCACCCCTATCCTGGCAGAATC **CG**GGCTTTGCCCTGGTATCAAGTCA **CG**GGAAGCTC **CG**GGAATGG **CG**GCCAGGCA **CG**T
 TCCC-----ATCCCCACCCACCCCTATCCTGGCAGAATC **CG**GGCTTTGCCCTGGTATCAAGTCA **CG**GGAAGCTC **CG**GGAATGG **CG**GCCAGGCA **CG**T

Figure S4. Nucleotide sequence identity between MAGEA3 and MAGEA6/CT-GABRA3 5'-regions. Genomic segments corresponding to the 5'-region of MAGEA3 (GRCh37 location reference: 151939225-151937726) and MAGEA6 (151866245-151867744) are aligned. Broken arrows indicate positions of the transcription start sites of MAGEA3 (ENST00000417212), MAGEA6 (ENST00000412733) and CT-GABRA3 (Genbank #KJ620007). Positions relative to MAGEA3 and MAGEA6 transcription start sites are indicated. Sequences are 100% identical, except for a 7bp deletion in intron 1 of MAGEA3. CpG methylation sites, highlighted in yellow, are perfectly conserved.