

Mechanism of Polycomb recruitment to CpG islands as revealed by inherited disease-associated mutation

Valentina S. Caputo¹, Joana R. Costa^{1,2}, Kalliopi Makarona¹, Elisabeth Georgiou^{1,3},
Mark Layton¹, Irene Roberts¹, Anastasios Karadimitris¹

¹Centre for Haematology, Department of Medicine, Imperial College London, Du Cane Road, London W12 0NN, UK

^{1,2}CEDOC, Faculdade de Ciências Médicas, FCM, Universidade Nova de Lisboa, 1169-056 Lisboa, Portugal

³Department of Biological Chemistry, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

Correspondence

Anastasios Karadimitris

Centre for Haematology

Imperial College London

London W12 0NN

e-mail: a.karadimitris@imperial.ac.uk

How the transcription repressing complex Polycomb interacts with transcriptional regulators at housekeeping genes in somatic cells is not well understood. By exploiting a CpG island (CGI) point mutation causing a Mendelian disease we show that DNA binding of activating transcription factor (TF) determines histone acetylation and nucleosomal depletion commensurate with Polycomb exclusion from the target promoter. Lack of TF binding leads to reversible transcriptional repression imposed by nucleosomal compaction and consolidated by Polycomb recruitment and establishment of bivalent chromatin status. Thus, within a functional hierarchy of transcriptional regulators, TF binding is the main determinant of Polycomb recruitment to the CGI of a housekeeping gene in somatic cells.

Introduction

Polycomb is a highly conserved, transcriptional repressive complex critical for transcriptional regulation of genes involved in development(1). Polycomb repressed genes are poised i.e., can be either activated or permanently silenced during developmental and cell differentiation processes. The functional hierarchy of *cis* and *trans* regulators of transcription that determines Polycomb recruitment to gene promoters of mammalian cells and in particular to the promoters of housekeeping genes in somatic cells remains to be established. To generate an integrated network of Polycomb and of *in cis* and *trans* regulators and its influence on transcription of active genes in somatic cells, we took advantage of the pathogenic C>G mutation in the promoter of the housekeeping gene *PIGM*(2). This mutation causes inherited glycosylphosphatidylinositol (GPI) deficiency (IGD), an autosomal recessive disorder characterized by epilepsy and pro-thrombotic diathesis(3). We compared TF binding, nucleosome dynamics, histone and DNA modifications in B cells from IGD patients (Mut) with those from normal (WT) and parental, i.e., heterozygous (Het) B cells. A distinct advantage of this approach is the ability to monitor the impact of changes in these processes on the surface expression of GPI in Mut cells as a reflection of *PIGM* mRNA and protein levels(3).

Results and Discussion

We previously showed that the pathogenic C>G mutation falls in a -270GC-rich box in the core promoter of *PIGM* thereby abolishing binding of TF Sp1 *in vitro*(2, 4). This GC-rich box is part of a 697bp CpG island spanning the promoter of *PIGM* (**Fig 1A**). Using chromatin immunoprecipitation in conjunction with real time PCR (ChIP-RQ-

PCR), we show binding of Sp1 to the -270GC-rich box *in vivo* in WT but not Mut B cells (**Fig 1B and Suppl Fig 1A**) while *PIGM* mRNA levels in Mut cells were <1% of normal indicating a strong repressing effect of the pathogenic mutation (**Fig 1C**). Although not tested directly, it is possible that other Sp factors such as Sp3 also bind to the -270GC-rich motif and may thus be implicated in the pathogenesis of IGD. We found histone (H) 3 and 4 hypoacetylation in the mutant *PIGM* promoter (**Fig 1D and Suppl Fig 1B**) concomitant with lack of Sp1 binding both of which are restored to normal levels upon exposure to HDAC inhibitors including butyrate(5) (**Fig 1E & F, Suppl Fig 1C&D**), thus implicating HDAC as important in gene repression in IGD. This is supported by the ability of butyrate to control treatment-intractable epilepsy in patients with IGD(4). H4 re-acetylation, increased Sp1 binding and an increase in *PIGM* mRNA levels in Mut B cells *in vitro* was observed at 30min, 1hr and 3hrs respectively, post-exposure to butyrate reaching levels comparable to those of WT cells by 24hrs (**Fig 1G and Suppl Fig 1C-E**), thereby resulting in restoration of the GPI biosynthetic pathway as assessed by expression of GPI at the cell surface (**Fig 1H and Suppl Fig 1F**). These effects of butyrate on the mutant *PIGM* promoter acetylation and transcription are specific as they are not observed in another housekeeping gene such as *GAPDH* (**Suppl Fig 1G**).

The increase in *PIGM* mRNA was not affected by the protein synthesis inhibitor cyclohexamide, suggesting a direct effect of butyrate on *PIGM* transcriptional activation (**Suppl Fig 1H**). Consistent with the notion that de-repression of *PIGM* transcription in response to butyrate requires restoration of Sp1 binding, transcriptional activation of *PIGM* was abrogated by mithramycin A (**Suppl Fig 1I**), an agent that inhibits binding of GC-rich motif-binding TF, including Sp1 (6-8). Together, these data show that Sp1 binding to the -270GC-rich motif is required for histone acetylation and transcriptional activation of *PIGM* and conversely, re-acetylation of histones is associated with restoration of Sp1 binding and *PIGM* expression.

Based on these observations, we first explored the role of CGI in nucleosome dynamics. CGI are promoter GC-rich DNA modules critical for transcriptional regulation of active genes(9, 10) and are reported to inherently de-stabilize nucleosome assembly at the promoters of housekeeping and active, tissue-specific genes, thus promoting RNA Polymerase II recruitment and gene transcription(10-12). Whether this process is independent of activating TF binding is unclear. By employing a restriction enzyme (RE) accessibility assay we found that the *PIGM*

promoter was 6-fold more 'accessible' in WT than Mut cells (**Fig 2A**) consistent with nucleosomal compaction in the latter. To gain more detailed insights into the impact of the C>G mutation on nucleosome positioning, we employed a micrococcal nuclease protection assay combined with RQ-PCR. We found low levels of nucleosome protection and thus nucleosome occupancy along the length of *PIGM* promoter in WT cells (**Fig 2B**), particularly in the area encompassing the transcription start site (TSS) and the -270GC-rich box, thus defining the nucleosome depleted region (NDR) (**Fig 2B and Suppl Fig 2A**). Significantly higher levels of protection across the NDR were seen in Mut than WT cells (**Fig 2B**), consistent with a more compact nucleosome assembly and positioning in the Mut promoter. Restoration of Sp1 binding following butyrate treatment had a dramatic effect on nucleosome compaction across the mutant *PIGM* core promoter, with protection levels decreasing towards those found in the core promoter of *PIGM* and of the housekeeping gene *GAPDH* in WT cells, presumably as a result of histone re-acetylation, suggesting restoration of the NDR (**Fig 2B**).

Further analysis of the core promoter nucleosome positioning by ligation mediated-PCR revealed a pattern consistent with a 'moving' or 'unstable' nucleosome in WT cells characteristic of active promoters (**Fig 2C**). By contrast, in Mut cells findings were consistent with a rigid, 'fixed' nucleosome positioning and occlusion of NDR (**Fig 2C**). Upon treatment of Mut cells with butyrate, the 'fixed' pattern reverted to a pattern resembling that of the 'moving' nucleosomes.

Based on the sequence and length of the amplicon spanning the core promoter (**Fig 2D and Suppl Fig 2B**), we found that while in WT cells the -270GC-rich box resides within the NDR, in Mut cells this motif falls in the middle of the 146bp stretch of DNA wrapped around the NDR nucleosome, potentially preventing Sp1 from accessing its cognate, GC-rich motif (**Fig 2D**). Taken together, the strikingly different nucleosome configurations in the WT and mutant promoters show that CGI, at least in the case of *PIGM* promoter, are not sufficient by themselves to impose a NDR, a pre-requisite for efficient transcription. Instead, binding of the activating TF Sp1 to the -270GC-rich motif and the consequent histone acetylation appears to be a critical requirement for nucleosomal configuration that is compatible with active gene transcription.

In view of the dynamic nature of nucleosome compaction and taking into account the temporal relationship of the *in cis* and *trans* events shown in Fig 1E-G, we propose that the C>G mutation does not completely abolish binding of Sp1 to the mutated -270GC-rich motif but it reduces it sufficiently so that it may lead to reduced recruitment of histone acetyltransferases (HAT) by Sp1, decreased nucleosome acetylation and compaction and thus reduced accessibility of the -270GC-rich motif. These events can operate in a self-enforcing, co-operative manner until the mutant motif is inaccessible to Sp1 most of the time. This model of self-enforcement may unfold in reverse in response to HDAC inhibitors: increased histone acetylation leading to nucleosome relaxation and increased accessibility of the -270GC-rich motif, even in its mutant form, allows some Sp1 binding. This leads to Sp1-dependent recruitment of HAT, promoting further histone acetylation, nucleosome relaxation and motif accessibility and eventually full restoration of Sp1 binding and transcriptional activation.

Since transcriptional repression by nucleosome compaction can be further consolidated by changes in DNA and histone methylation(13), we analyzed the mutant promoter for the repressive H3K9me2 and H3K27me3 histone modifications, characteristic marks of heterochromatin and developmentally regulated/poised genes respectively(14). While there was no difference in H3K9me2 levels (**Fig 3A**), compared to WT, we observed a significant enrichment for the H3K27me3 mark in the mutant promoter (**Fig 3B and Suppl Fig 3A & B**), consistent with transcriptional repression of *PIGM* by Polycomb (PcG)(1, 15, 16). Supporting this, Suz12, EzH2 and Ring1B, components of the PcG PRC2 and PRC1 complexes respectively(1), were significantly enriched in the mutant promoter as was Jarid2, a de-methylase that contributes to the recruitment of PcG to GC-rich domains(17, 18) (**Fig 3C and D**). Treatment of Mut cells with DZNep, a pharmacological agent that depletes PRC2 components(19-21), including EzH2, increased *PIGM* mRNA levels by up to 70-fold (**Fig 3E**), supporting the importance of PcG-mediated transcriptional repression of the mutant promoter. Further, restoration of Sp1 binding in response to butyrate caused a time-dependent decrease in the enrichment of EzH2 and H3K27me3 but not of H3K4me3 marks in the mutant *PIGM* promoter (**Fig 3F**; see below). Thus, PcG recruitment and deposition of PcG-mediated H3K27me3 marks to the mutant CGI consolidates the nucleosome compaction caused by the C>G mutation and the lack

of Sp1 binding to the core promoter -270GC-rich motif. These data also conclusively show that lack of activating TF binding and gene transcription at GC-rich promoters leads to PcG recruitment.

To analyze the impact of the C>G mutation on the activating histone mark status we performed ChIP-RQ-PCR for the H3K4me3 modification which, as expected for a housekeeping gene(14), was found highly enriched in the WT *PIGM* gene (**Fig 4A**) and in particular in the core promoter (**Suppl Fig 4A**). Surprisingly, we also found comparable or higher than in WT H3K4me3 enrichment in the mutant promoter (**Fig 4A**) suggesting the presence of a bivalent chromatin state of the antagonistic H3K4me3/H3K27me3 marks. Co-existence of these marks in the promoters of developmental genes in embryonic stem cells is required for their dynamic regulation during development and differentiation(22-25). To establish that bivalent status represents true bivalency on the same promoter, we studied histone methylation in parental heterozygous B cells (**Suppl Fig 4B**). Taking advantage of the novel BstUI RE site created by the C>G mutation (**Fig 1A and Suppl Fig 4B**) we found that upon BstUI digestion in Het cells, levels of enrichment of H3K4me3 and CFP1, a CGI-binding protein required for H3K4 trimethylation(26), decreased by ~50%, (**Fig 4B & C**) while enrichment of H3K27me3, EzH2, Ring1B and Suz12 decreased by >70% (**Fig 4D & E and Suppl Fig 4C**). These results are consistent with simultaneous presence at the Mut *PIGM* promoter of H3K4me3/H3K27me3 marks and PcG complexes. We further confirmed co-occupancy by H3K4me3/H3K27me3 marks on the same mutant but not WT *PIGM* promoter using reciprocal re-ChIP assays in conjunction with BstUI digestionⁱ in Het cells (**Fig 4F & G**). These findings indicate that in contrast to histone acetylation which is dependent on Sp1 binding, bivalent histone methylation marks and PcG recruitment to CGI, are default processes instructed by the absence of activating TF binding and dissociated from active transcription, thus confirming and extending previous reports(16, 27-29).

Finally, bisulphite sequencing showed that the *PIGM* CGI is unmethylated in WT cells (**Fig 4G**), consistent with active gene status(30). DNA methylation status of the *PIGM* CGI in Mut cells did not change and there was no change in *PIGM* mRNA levels and surface GPI expression in these cells when they were treated with the DNA de-methylating agent 5'-azacytidine (**Suppl Fig 4D**). These findings are contrary to earlier as well as more recent data that directly or indirectly link lack of TF, including of Sp1 binding to cognate motifs to methylation of regulatory DNA(31-33). Instead, they indicate that lack of activating TF binding or PcG-mediated

transcriptional repression are not by themselves sufficient for CGI DNA methylation and gene silencing. Instead, they are in line with genome-wide and structural data(34, 35) showing that presence of H3K4me3 marks protects CGI promoters from DNA methylation and thus permanent transcriptional silencing. They are also consistent with previous work showing that in Sp1-/- murine embryonic stem cells, CGI of genes regulated by Sp1 are not methylated(36).

In conclusion, our data strongly support a model of a functional hierarchy, which, with activating TF as its central hub, determines recruitment or exclusion of PcG from CGI (**Suppl Fig 5**). Specifically, we propose that histone acetylation and nucleosomal depletion, PcG exclusion and thus active transcription are all dependent on activating TF binding to CGI. By contrast, lack of TF binding histone is associated with conservation of activating H4K3me3 histone marks, PcG recruitment and imposition of repressing H3K27me3 methylation and thus conversion of an active promoter to a repressed, bivalent chromatin state. Preservation of H4K3me3 histone marks at CGI, even at the repressed state, prevents DNA methylation by DNA methyltransferases, possibly through steric hindrance(35), thus averting permanent gene silencing. In addition, our observations constitute the first example of PcG-mediated bivalent chromatin domain transcriptional repression as a pathogenetic mechanism in Mendelian disease.

Acknowledgements

This work was supported by the Medical Research Council (VC, ML, AK), the Fundacao para a Ciencia e Tecnologia, Lisboa, Portugal (JRC), the Foundation of the Hellenic Society of Haematology (EG), the Halley Stewart Trust (KM and IR) and the NIHR Biomedical Research Centre at Imperial College Healthcare NHS Trust.

We thank Ana Pombo, Doug Higgs and members of our laboratory for critical reading of the manuscript.

Materials and Methods

Cell lines and treatment with inhibitors

Normal (WT), patient (Mut) and parental (Het) EBV-immortalised lymphoblastoid B cells were cultured in RPMI1640 with 10% FBS, 2mM L-glutamine and 1 unit/ml penicillin- 1µg/ml streptomycin. Mut1, Mut2, Het1, Het2 and Het3 are from members of the same family while Mut3 are cells from a patient from a different family.

Sodium butyrate (Bu), Mithramycin A (MitA), 5'-azacytidine (5'-Aza), Cyclohexamide (CHX) and Trichostatin A (TSA) were from Sigma. SAHA (Selleck Chemicals),

DzNep (Cambridge Bioscience). Bu and 5'-Aza were prepared fresh in RPMI complete medium at 600mM and 5mM stocks respectively. MitA, CHX, TSA, SAHA and DzNep were dissolved in DMSO at 1mM, 100mg/ml, 50 μ M, 200mM, 20mM stock solutions and diluted prior to use at the indicated final concentrations.

Gene expression analysis

Total RNA was extracted using the GeneJet RNA purification kit (Fermentas-THERMO). After DNA traces were removed by DNase treatment (Fermentas-THERMO), cDNA was synthesised with random primers using the ReverAid First strand cDNA Synthesis Kit (Fermentas-THERMO). RQ-PCR experiments were carried out on AB 7500 Real time PCR system using Maxima SYBR Green (Fermentas-THERMO). The primers used are listed in **Suppl Table 1**.

Chromatin Immunoprecipitation

This was performed using a described in ChIPAb+Sp1 Millipore kit with some modifications. Specifically, cells were collected by centrifugation at 1200 rpm. Chromatin was prepared using the Shearing ChIP kit (Diagenode). Chromatin was sonicated at 4⁰C for 20 minutes in a 0.5/0.5 on/off cycle using high intensity in a Bioruptor UCD-200 (Diagenode).

Post-sonication fragments of average 500bp length were confirmed on a 1.5% agarose gel. The chromatin was diluted at least 10 times in ChIP dilution buffer (0,01%SDS, 1,1% Triton X-100, 1,2mM EDTA, 16,7mM Tris-HCL pH 8, 167mM NaCl) containing protein inhibitors. The diluted chromatin was pre-cleared by incubation with magnetic beads (Dynabeads Protein G form Invitrogen) for 1hr at 4⁰C on a rotating wheel. Pre-cleared chromatin was incubated with 3 μ g of antibodies overnight at 4⁰C on a rotating wheel.

Protein G magnetic beads were added and incubated for 2-4hrs at 4⁰C on a rotating wheel. The immunoprecipitates were washed for 5 min at 4⁰C in a rotating wheel with low salt buffer (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl pH 8, 150mM NaCl), high salt buffer (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl pH 8, 500mM NaCl), LiCl buffer (0,25M LiCl, 1 % IGEPAL, 1% sodium deoxicolate, 1mM EDTA, 10mM Tris-HCl pH 8) and 2 times with TE.

Immunocomplexes were eluted by adding 150 μ l elution buffer (50mM Tris-HCl pH 8, 50mM NaCl, 1mM EDTA and freshly added 1% SDS and 20mg/ml of RNaseA) at 65⁰C for 4h to overnight and a second time for 30 min. Both elutions were pooled and treated with proteinase K (Fermentas-THERMO).

DNA was purified by phenol extraction and DNA precipitation. RQ-PCR was performed using Maxima SYBR Green (Fermentas-THERMO) on an AB 7500 Real time PCR system.

Enrichment of the target sequence was assessed against an isotypic IgG antibody - immunoprecipitated aliquot. The $\Delta\Delta Ct$ method was used to calculate the enrichment of the target sequence, compared to the negative IgG control: $\Delta\Delta Ct = \Delta Ct (Ct_{\text{sample}} - Ct_{\text{input}}) - \Delta Ct (Ct_{\text{IgG}} - Ct_{\text{input}})$. The fold difference of the experimental sample relative to the IgG control (both normalized to the input sample) was expressed by the power $2^{-\Delta\Delta Ct}$.

Allele specific ChIP assay

ChIP from heterozygous B cell lines was performed as described above. After purification, equal amounts of DNA were diluted in restriction enzyme buffer and treated or not with BstUI under the same conditions. Under these conditions, while the *PIGM* mutant promoter allele is sensitive to BstUI digestion (**Fig 1a**), the undigested DNA corresponds to both, mutated and unmutated alleles. Digestion efficiency was assessed by qPCR using a BstUI-sensitive *GAPDH* promoter amplicon as reference and was found to be >95% in all cases.

Reference List

1. Simon JA, Kingston RE. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol.* 2009 10/2009;10(10):697-708.
2. Almeida AM, Murakami Y, Layton DM, Hillmen P, Sellick GS, Maeda Y, et al. Hypomorphic promoter mutation in PIGM causes inherited glycosylphosphatidylinositol deficiency. *NatMed.* 2006 7/2006;12(7):846-51.
3. Almeida A, Layton M, Karadimitris A. Inherited glycosylphosphatidyl inositol deficiency: a treatable CDG. *BiochimBiophysActa.* 2009 9/2009;1792(9):874-80.
4. Almeida AM, Murakami Y, Baker A, Maeda Y, Roberts IA, Kinoshita T, et al. Targeted therapy for inherited GPI deficiency. *N Engl J Med.* 2007 4/19/2007;356(16):1641-7.
5. Davie JR. Inhibition of histone deacetylase activity by butyrate. *J Nutr.* 2003 7/2003;133(7 Suppl):2485S-93S.
6. Malek A, N+|+|ez LE, Magistri M, Brambilla L, Jovic S, Carbone GM, et al. Modulation of the Activity of Sp Transcription Factors by Mithramycin Analogues as a New Strategy for Treatment of Metastatic Prostate Cancer. *PLoS ONE.* 2012 4/19/2012;7(4):e35130.
7. Blume SW, Snyder RC, Ray R, Thomas S, Koller CA, Miller DM. Mithramycin inhibits SP1 binding and selectively inhibits transcriptional activity of the dihydrofolate reductase gene in vitro and in vivo. *J Clin Invest.* 1991 11/1991;88(5):1613-21.
8. Barcel+| F, Scotta C, Ortiz-Lombard+|a M, M+@ndez C, Salas JA, Portugal J. Entropically-driven binding of mithramycin in the minor groove of C/G-rich DNA sequences. *Nucleic Acids Research.* 2007 4/1/2007;35(7):2215-26.
9. Blackledge NP, Klose R. CpG island chromatin: a platform for gene regulation. *Epigenetics.* 2011 2/2011;6(2):147-52.
10. Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes Dev.* 2011 5/15/2011;25(10):1010-22.
11. Ramirez-Carrozzi VR, Braas D, Bhatt DM, Cheng CS, Hong C, Doty KR, et al. A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell.* 2009 7/10/2009;138(1):114-28.
12. Hargreaves DC, Horng T, Medzhitov R. Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell.* 2009 7/10/2009;138(1):129-45.
13. Beisel C, Paro R. Silencing chromatin: comparing modes and mechanisms. *Nat Rev Genet.* 2011 2/2011;12(2):123-35.
14. Kouzarides T. Chromatin modifications and their function. *Cell.* 2007 2/23/2007;128(4):693-705.
15. Bracken AP, Helin K. Polycomb group proteins: navigators of lineage pathways led astray in cancer. *Nat Rev Cancer.* 2009 11/2009;9(11):773-84.
16. Guenther MG, Young RA. Transcription. Repressive transcription. *Science.* 2010 7/9/2010;329(5988):150-1.
17. Landeira D, Sauer S, Poot R, Dvorkina M, Mazzearella L, Jorgensen HF, et al. Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nat Cell Biol.* 2010 6/2010 print;12(6):618-24.
18. Shen X, Kim W, Fujiwara Y, Simon MD, Liu Y, Mysliwiec MR, et al. Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell.* 2009 12/24/2009;139(7):1303-14.
19. Miranda TB, Cortez CC, Yoo CB, Liang G, Abe M, Kelly TK, et al. DZNep is a global histone methylation inhibitor that reactivates developmental genes not silenced by DNA methylation. *Mol Cancer Ther.* 2009 6/2009;8(6):1579-88.
20. Xie Z, Bi C, Cheong LL, Liu SC, Huang G, Zhou J, et al. Determinants of Sensitivity to DZNep Induced Apoptosis in Multiple Myeloma Cells. *PLoS ONE.* 2011 6/24/2011;6(6):e21583.

21. Zhang X, Zhao X, Fiskus W, Lin J, Lwin T, Rao R, et al. Coordinated Silencing of MYC-Mediated miR-29 by HDAC3 and EZH2 as a Therapeutic Target of Histone Modification in Aggressive B-Cell Lymphomas. *Cancer Cell*. 2012 10/16/2012;22(4):506-23.
22. Bernstein BE, Meissner A, Lander ES. The mammalian epigenome. *Cell*. 2007 2/23/2007;128(4):669-81.
23. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006 4/21/2006;125(2):315-26.
24. Azuara V, Perry P, Sauer S, Spivakov M, Jorgensen HF, John RM, et al. Chromatin signatures of pluripotent cell lines. *Nat Cell Biol*. 2006 5/2006;8(5):532-8.
25. Fisher CL, Fisher AG. Chromatin states in pluripotent, differentiated, and reprogrammed cells. *Curr Opin Genet Dev*. 2011 4/2011;21(2):140-6.
26. Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, et al. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*. 2010 4/15/2010;464(7291):1082-6.
27. Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, et al. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet*. 2008 10/2008;4(10):e1000242.
28. Lynch MD, Smith AJ, De GM, Flenley M, Hughes JR, Vernimmen D, et al. An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *EMBO J*. 2012 1/18/2012;31(2):317-29.
29. Mendenhall EM, Koche RP, Truong T, Zhou VW, Issac B, Chi AS, et al. GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet*. 2010 2010;6(12):e1001244.
30. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*. 2012 7/2012;13(7):484-92.
31. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. *Nature*. 2012 9/6/2012;489(7414):75-82.
32. Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, Nemes A, et al. Sp1 elements protect a CpG island from de novo methylation. *Nature*. 1994 Sep 29;371(6496):435-8. PubMed PMID: 8090226.
33. Macleod D, Charlton J, Mullins J, Bird AP. Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes & development*. 1994 Oct 1;8(19):2282-92. PubMed PMID: 7958895.
34. Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet*. 2007 4/2007;39(4):457-66.
35. Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*. 2007 8/9/2007;448(7154):714-7.
36. Marin M, Karis A, Visser P, Grosveld F, Philipsen S. Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. *Cell*. 1997 May 16;89(4):619-28. PubMed PMID: 9160753.

Figure legends

Figure 1. Sp1-dependent reversible transcriptional repression associated with the *PIGM* core promoter -270GC-rich box C>G mutation

(A) Schematic representation of *PIGM*, a single exon gene. Note that -270 refers to distance from the start codon ATG. **(B)** ChIP analysis of Sp1 relative enrichment at the *PIGM* promoter of WT and Mut B cells, with a *DHFR* promoter region used as a positive control (n=3). **(C)** *PIGM* mRNA levels relative to *GAPDH* in WT, Het and Mut B cells (n=3). **(D)** ChIP analysis for histone 3 (H3Ac) and histone 4 (H4Ac) acetylation enrichment at the *PIGM* promoter of WT and Mut B cells (n=3). **(E-G)** ChIP analysis of Sp1(**E**) and H4Ac enrichment (**F**) at the *PIGM* promoter and *PIGM* mRNA levels (**G**) in response to butyrate (Bu; 3mM) treatment over 36 hrs. Representative of 3 independent experiments. Data are normalized to WT 0hrs values. **(H)** FACS analysis of cell surface GPI expression after staining with FLAER of Mut, WT and Het B cells (left panel) and of Mut B cells upon Bu treatment (right panel) for the time points shown. All data are shown as mean $\pm$ SEM.

Figure 2. Nucleosomal positioning and occupancy associated with lack of Sp1 binding to CGI

(A) Restriction enzyme accessibility assay in WT and Mut B cells. *AatII* digestion was performed in isolated nuclei to monitor accessibility within the core promoter of *PIGM*. Digestion levels were quantified by RQ-PCR relative to a control region of the *PIGM* promoter lacking *AatII* restriction site (amplicons 2 and 5 in Fig 1A; n=3). **(B)** Nucleosome mapping by MNase protection assay on WT, Mut and Bu-treated (3mM, 15hrs) Mut B cells. The abundance of twelve 100bp amplicons that overlap by 50bp covering the length of the *PIGM* promoter region was quantified within equal amounts of MNase-digested or undigested DNA by RQ-PCR. *GAPDH* promoter amplicon is shown as an accessibility control for WT, Mut and butyrate-treated B cells. Plots show mean levels of relative protection $\pm$ SEM (n=3). Amplicon b is shown in bold. **(C)** Ligation-mediated PCR analysis in WT and Mut B cells with and without Bu treatment using an amplicon b forward primer and an adaptor-specific reverse primer. **(D)** *PIGM* promoter DNA sequence. *AatII* and *BstUI* sites are underlined. C>G mutation is shown with a star, -270 Sp1 binding site is shown in

italics. Nucleotide sequence of the promoter DNA corresponding to the nucleosome spanning the TSS (arrow) and -270GC rich box in the Mut promoter is highlighted in grey (see also Supplemental Fig 2B).

Figure 3. Lack of Sp1 binding to CGI results in Polycomb-recruitment and repressive H3K27me3 mark deposition

(A-D) ChIP analysis of H3K9me2 (A), H3K27me3 (B), Jarid2, Suz12 (C), EzH2 and Ring1B (D) enrichment at the *PIGM* promoters of WT and Mut B cells (n=3). HOXA2, a member of the Hox genes that are known to be repressed by PcG and H3K27 trimethylation is shown as positive control. **(E)** *PIGM* mRNA levels in Mut, WT and Het B cells after treatment for 72hrs with the EzH2 inhibitor DzNep (3 μ M). Data normalised to DMSO-treated cells (n=3). **(F)** Enrichment of H3K27me3, H3K4me3, and EzH2 at the *PIGM* promoter of Mut B cells after 5 and 24 hrs treatment with Bu (3mM; n=3). All data are shown as mean $\pm$ SEM.

Figure 4. Absence of Sp1 binding to CGI leads to bivalent chromatin domain-mediated transcriptional repression but not to DNA methylation

(A) ChIP analysis of H3K4me3 enrichment at the *PIGM* promoter of WT and Mut B cells (n=3). **(B-E)** Allele-specific ChIP analysis (see Supplemental Fig 4B) in Het1 cells. ChIP analysis of H3K4m3, CFP1, H3K27m3, EzH2, Suz12 and Ring1B enrichment levels at the *PIGM* promoter of Het B cells with and without BstUI digestion (see Fig 1A). Enrichment for H3K27me3, EzH2, Suz12 and Ring1B at an area of *HOXA2* gene that lacks BstUI restriction site is shown as a positive control (n=3). **(F-G)** Allele specific, reciprocal re-ChIP analysis of H3K27me3-H3K4me3 co-occupancy in Het1 B cells. Large graphs show enrichment in the second ChIP, with and without BstUI digestion, normalized to IgG control. Insets show enrichment as % of Input in the first ChIP of the re-ChIP assay. All data are shown as mean $\pm$ SEM (n=3). **(H)** *PIGM* CGI bisulphite sequencing of two different Mut and one WT B cell line. Filled circles represent methylated CpG, novel CpG created by the C>G mutation is indicated by and arrowhead. Transcription start site (TSS) is shown by an arrow, % methylation is shown on the right. Methylation at the Oct4 gene is shown as a positive control (*right*)
