Pax7 pioneer factor action requires both paired and homeo DNA binding domains

Audrey Pelletier[1,3], Alexandre Mayran[1,2], Arthur Gouhier[1], James G. Omichinski [ID][3], **Aurelio Balsalobre[1] and Jacques Drouin** [ID][1,3,*]

[1]Laboratory of Molecular Genetics, Institut de recherches cliniques de Montréal (IRCM), Montréal QC H2W 1R7, Canada, [2]EPFL SV ISREC UPDUB, CH-1015 Lausanne, Switzerland and [3]Department of Biochemistry, Faculté of Médecine, Université de Montréal, Montréal H3C 3J7, Canada

ABSTRACT

The pioneer transcription factor Pax7 contains two DNA binding domains (DBD), a paired and a homeo domain. Previous work on Pax7 and the related Pax3 showed that each DBD binds a cognate DNA sequence, thus defining two targets of binding and possibly modalities of action. Genomic targets of Pax7 pioneer action leading to chromatin opening are enriched for composite DNA target sites containing juxtaposed sites for both paired and homeo domains. The present work investigated the implication of the DBDs in pioneer action. We show that the composite sequence is a higher affinity binding site and that efficient binding to this site involves both DBDs of the same Pax7 molecule. This binding is not sensitive to cytosine methylation of the DNA sites consistent with pioneer action within nucleosomal heterochromatin. Introduction of single amino acid mutations in either paired or homeo domain that impair binding to cognate DNA sequences showed that both DBDs must be intact for pioneer action. In contrast, only the paired domain is required for low affinity binding of heterochromatin sites. Thus, Pax7 pioneer action on heterochromatin requires unique protein:DNA interactions that are more complex compared to its simpler DNA binding modalities at accessible enhancer target sites.

INTRODUCTION

Pioneer factors are a subset of transcription factors (TFs) that have been dubbed with the unique property of being able to 'open chromatin structure' (1–3). While pioneers have the typical TF domain organisation found in most TFs including DNA binding and transactivation domains, the molecular basis of their ability to 'open chromatin' remains largely elusive. In fact, recent work clearly separated the chromatin opening function from the pioneer ability and showed that the former can be conferred through recruitment of a nonpioneer factor: this was shown for the pioneer Pax7 that is required to implement the melanotrope enhancer repertoire (4,5) but that is dependent on the cell determination TF Tpit for chromatin opening (6). These findings thus limit the unique pioneer property to recognition of DNA binding sites within heterochromatin. A large body of work also showed that pioneer TFs can bind their target DNA sequence within nucleosomes that would presumably be tightly packed within heterochromatin (7,8). Whereas the ability to recognize target DNA within nucleosomal structures appears to be a prerequisite for pioneer action, this in itself is not sufficient to explain the specificity of pioneer action nor to explain how initial sequence recognition is linked to chromatin remodelling.

Various pioneers activate novel enhancer repertoires that control cell specific gene expression programs. This was shown for the pioneers FoxA (9,10), Pax7 (4,5), C/EBPα (11), EBF1 (12,13), Ascl1 (14), NeuroD1 (15) and GATA3 (16). For all these factors, with the exception of Pax7, the DNA binding sites for the TF at pioneered enhancers (hereafter, we refer to 'pioneered enhancers/sites' as those that require chromatin opening for activity) appeared similar to their previously defined transcriptional target sites. In contrast, the initial work on Pax7 pioneer target sites revealed enrichment for a unique composite target sequence that includes motifs recognized by each DNA binding domain (DBD) present in Pax7 (5). Indeed, Pax7 contains two DBDs, a paired domain (PD) and a homeo (HD) domain, and each recognizes different prototypical DNA target sites. Prior work on the DNA binding properties of Pax7 and the closely related Pax3 used cognate binding sites for either the PD or the HD (17), and these appear to show independent binding by the PD or HD domains. However, close interaction of the two DBDs appears important for optimal DNA binding (18,19). Indeed, DNA binding by either the PD or HD causes a structural change in the other DBD (20). In this context, recognition of a unique compos-

*To whom correspondence should be addressed. Tel: +1 514 987 5680; Email: jacques.drouin@ircm.qc.ca

ite DNA sequence containing binding motifs for both PD and HD enriched at pioneered targets suggest that Pax7 pioneer action may involve a unique conformation or binding requirement.

It is presently unclear what DNA recognition features may be critical for pioneer target recognition and action. Although higher affinity for DNA sites present at pioneer targets may be a mechanism to trigger structural changes at nucleosomal sites, various studies have shown the ability of pioneer TFs to recognize subsets of their target sequences by recognition of partial and presumably lower affinity DNA motifs. This was proposed to represent a 'scanning' ability of pioneers to probe putative DNA targets within nucleosomal heterochromatin (21). This feature results in large numbers of pioneer-bound sites that are unproductive for chromatin remodelling and were thus labelled as 'resistant' (4) or as unchanged (21,22). The contribution of DNA binding affinity for pioneer action thus remains unclear. It was recently shown that the FoxA pioneer factor requires a core histone interaction domain for chromatin opening (23). However, the scanning ability and core histone interaction do not explain the selection of site subsets that are productive for chromatin opening.

The present work investigated the DNA binding properties of Pax7 in relation to its various target DNA sequences both *in vitro* using gel mobility shift and *in vivo* using ChIPseq. The highest *in vitro* binding affinity was observed with a composite Pax7 target that includes DNA motifs for recognition by both PD and HD and this composite sequence is enriched at pioneered enhancers. The composite DNA site is bound *in vitro* by a Pax7 monomer and the binding affinity is unchanged when the composite sequence contains a methylated CpG, a feature that would be expected of pioneers targeting regulatory sequences within nucleosomal heterochromatin. Finally using single amino acid substitutions of key residues required for DNA binding of either PD or HD, we show that the DNA binding activities of both domains are required for pioneer action whereas the PD alone is sufficient for binding to low affinity resistant sites. In summary and in contrast with the ability of each DNA binding domain to bind its target sequence in naked DNA, the dependence on both domains in order to trigger chromatin opening at pioneered sites suggests that a unique conformation of the Pax7 DBDs is needed for the initial events leading to chromatin opening.

MATERIALS AND METHODS

Cell culture

AtT-20/D16v-F2 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and antibiotics (penicillin/streptomycin). To generate stable transgenic AtT-20 Pax7 G43A, R56L and S264E cell populations, retroviruses were packaged using the Platinum-E Retroviral Packaging Cell Line (Cell Biolabs, catalog #RV-101) and infections performed as described (24). Selection of retrovirus-infected cell populations was achieved with 400 μg/mL Geneticin (Gibco, #11811-031). Resistant colonies were pooled to generate populations of hundreds of independent colonies.

Plasmids and luciferase assays

The PC2 reporter plasmid (5) and construction of oligonucleotide-driven reporters were described previously (25). Mutagenesis of mouse Pax7 coding sequences was carried out using the Stratagene QuickChange Site-Directed Mutagenesis kit as reported previously (26). AtT20 cells were plated in 12-well dishes (4×10^5 cells per well) and transfections performed the next day using PEI 25K (Polysciences #23966-1) in triplicates. Transfected DNA (total 2 μg in 150 μl DMEM) was added to the PEI 25K solution (4 μl per 150 μl DMEM) and incubated at room temperature for 30 minutes before adding 100 μl of the mixture to each well. The media was changed 24 h after transfection and luciferase activity assessed 48 hours after transfection. Growth medium was removed from transfected cells and 200 μl of lysis buffer (Tris 100 mM, NP40 0.5% and DTT 0.001 M) added to each well. After 10 min on a vigorous shaker, 100 μl of cell lysate supernatant was used for analysis using the GloMax Navigator luminometer (Promega) using the kinetics protocol (100 μl injection, speed 100 μl/s, integration time 10 s). Statistical significance was assessed by two-way ANOVA and Tukey`s multiple combinations test.

ChIPseq, ATACseq and MNaseSeq

ChIP experiments were performed in AtT-20 cells as previously described (4,27) using 5 μg of antibody (anti-FLAG-M2, mouse monoclonal, Sigma F3165). ATACseq was performed as previously described (4). The libraries and flow cells were prepared by the IRCM Molecular Biology Core Facility following Illumina recommendations (Illumina, San Diego, CA, USA). The ChIP libraries were sequenced on the Illumina Hi-Seq 2000 sequencer. Supplementary Table S2 provides details of antibodies used for ChIP and sequencing depths. Relative Pax7 ChIP and ATACseq read density profiles were generated by Easeq average signal function (28) (http://easeq.net) and data were exported to Microsoft Excel to create graphs. ChIPseq and ATACseq heatmaps were generated by Easeq. To visualize ChIPseq and ATACseq profiles, we used the Integrative Genome Viewer tool (29).

MNase-seq was performed as described (30) with the following modifications. AtT20 Neo and Pax7 cells (60 million) were lysed in 1 ml lysis buffer (10 mM Tris, pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.5% NP-40, 0.15 mM spermine, 0.5 mM spermidine, 1 mM PMSF, 0.6 μM aprotinin, 0.6 μM leupeptin and 2 μM pepstatin A). Nuclei were resuspended in 1 ml MNase digestion buffer (10 mM Tris–HCl pH 7.4, 15 mM NaCl, 60 mM KCl, 0.15 mM spermine, 0.5 mM spermidine, 1.5 mM $CaCl_2$, 1 mM PMSF, 0.6 μM aprotinin, 0.6 μM leupeptin and 2 μM pepstatin A). 50 μl of nuclei suspension (∼50 μg) were digested in 16 U MNase (Micrococcal Nuclease NEB M0247S) at 37 °C for 60 min. The reaction was stopped with 5 μl EDTA 250 mM and to break the nuclei, 5 μl SDS 10% was added. The volume was adjusted to a total of 200 μl with140 μl TE and 20 μl NaCl 5 M was added to dissociate histones from DNA. After phenol/chloroform extraction and ethanol precipitation, 2 μl of RNase (10 mg/ml) was added to the mixture and incubated at 37 °C for 30 min. DNA was purified using

Qiagen PCR purification kit and eluted with 30 μl EB. Sequencing libraries were performed as previously described (4).

Expression and purification of Pax7 polypeptides from bacteria

For the expression of polypeptide fragments of Pax7 in bacteria, the sequences encoding the PD alone (PD; residues 35–158 of mouse Pax7 Uniprot P47239-1),the HD alone (HD; residues 215–273 of mouse Pax7), the PD plus the linker region (PD-L; residues 35–215 of mouse Pax7) or the PD plus linker plus HD (PD-HD; residues 35–273 of mouse Pax7) were incorporated into the pET15b (Novagen) vector. The vectors were then transformed in *Escherichia. coli* host strain BL21DE3 (Novagen) and the cells were grown at 37°C in Luria Broth (LB) media overnight, diluted 1:4 with fresh LB and expression of the His-tagged Pax-7 polypeptides was induced for 4 h at 30°C with 0.7 mM isopropyl-β-D-thiogalactopyranoside (IPTG; Inalco). The cells were harvested by centrifugation, suspended in lysis buffer (either 20 mM Tris–HCl pH 7.4, 1M NaCl and 50 mM imidazole for HD, PD-L and PD-HD or 20 mM NaPO$_4$ pH 7.4, 1M NaCl and 50 mM imidazole for PD), lysed with a French press and centrifuged at 105 000g for 1h at 4°C. The supernatant was collected and incubated with a nickel charged Chelating-Sepharose resin (GE healthcare) for 1 h at 4°C. Following incubation, the resin was thoroughly washed with 20 mM Tris–HCl pH 7.4, 1 M NaCl and 50 mM imidazole, and the polypeptides wer eluted from the resin with elution buffer (20 mM Tris–HCl pH 7.4, 1 M NaCl and 500 mM imidazole). The eluted proteins were dialyzed against 20 mM Tris–HCl pH 7.4, 1mM EDTA buffer overnight and further purified on a 50 ml High Performance SP-Sepharose (GE Healthcare) column using a gradient with 20 mM Tris–HCl pH 7.4, 0.75 M NaCl,1mM EDTA over 5 column volumes. The fractions containing the purified polypeptides were dialyzed against 20 mM Tris–HCl pH 7.4, 1mM EDTA buffer overnight and stored at –80°C until usage.

Electrophoretic mobility shift assays (EMSA)

Probe labeling. Double-stranded DNA probes were prepared by annealing oligonucleotides (1 μM of each strand: see sequences in the Supplementary Figure S1A) in buffer TEN (10 mM Tris pH 8.0, 1 mM EDTA and 50 mM NaCl). Labeling was done at room temperature for 20 min in the following buffer: 50 mM Tris pH 7,5, 10 mM MgCl$_2$, 1 mM DTT and 0.5 mg/ml BSA supplemented with 0.25 μM dGTP, dTTP, 0,2 μM α-^{32}P-dCTP (3,000 Ci/mmol, 10 mCi/mL, Perkin Elmer) and 1 unit of DNA polymerase I Large (Klenow) fragment from *E. coli* (NEB, catalog # M0210M). Probes were then purified on Micro Bio-Spin *P*-30 Tris chromatography column (Bio-rad, catalog # 732-6223).

Nuclear extracts. To prepare nuclear extracts, AtT-20 cells from a 100mm petri dish were washed and collected in PBS. Pelleted cells were resuspended in 800 μl of cold low-salt buffer (10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM EDTA,

0.1 mM EGTA, 1 mM DTT, 0.5 mM PMSF, 0.6 μM aprotinin, 0.6 μM leupeptin and 2 μM pepstatin A). After 15 min incubation on ice, 100 μl NP40 10% was added. Cytosolic fractions were discarded, and nuclear proteins extracted in 100 μl of cold high-salt buffer containing 20 mM HEPES pH 7.9, 400 mM KCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 1 mM PMSF, 0.6 μM aprotinin, 0.6 μM leupeptin and 2 μM pepstatin A. Protein concentration was quantitated by Bradford assay.

EMSA. Gel shift assays were performed using 1 μg of nuclear protein extracts from different AtT-20 cell lines preincubated 10 min on ice with 1 μg of nonspecific carrier DNA poly(dI-dC) and poly(dA-dT) in binding buffer (25 mM HEPES pH 7.9, 84 mM KCl, 10% glycerol, 5 mM DTT). For purified polypeptides, 300 fmol were used per binding reactions. If applicable, cold competitor probe was added to the pre-incubation mix. Equal amounts of radiolabeled DNA probes (50 000 cpm) were then added for a total volume of 20 μl and incubated 60 min on ice. If applicable, 1 μg of Flag M2 antibody (F3165, SIGMA) was added to the mix during the last 10 min of incubation.

Electrophoresis. Nondenaturing polyacrylamide gels (40 mM Tris, 195 mM glycine, 0.08% APS, 0.5 μl/ml TEMED, 5% acrylamide:bisacrylamide in 19:1 ratio), were prerun in Tris–glycine buffer (40 mM Tris and 195 mM glycine) for 60 min at 300 V at 4°C. The gel shift reaction mixture was then loaded, and the gel run for 3 h at 300 V at 4°C. Gels were dried on 3MM Whatman paper in a gel dryer at 80°C during 60 min under vacuum. Autoradiography films (HyBlot CL, catalog # E3018) were then exposed with the dried gels for 10–48 h using an intensifier screen.

RT-qPCR

RNA was extracted from AtT-20 cells using the RNeasy Mini kit (Qiagen #74104), and cDNA was synthesized using 5 μg RNA and SuperScript III reverse transcriptase (Invitrogen, #18080-044) following manufacturer's recommendations. The resulting cDNAs (5 μl of 1/50 dilution) were analyzed by qPCR using SYBR Green reagent (ThermoFisher Scientific #A25741) supplemented with 500 nM of each gene-specific primer pair (provided in Supplementary Table S1) in a total volume of 10 μl. At least three biological replicates were analyzed in duplicates for each condition using a ViiA™ 7 Real-Time PCR device (ThermoFisher scientific), and results were analyzed using the accompanying software. Gene expression values were normalized to those of GAPDH and statistical significance assessed by two-way ANOVA and Tukey`s multiple combinations test.

NGS data analysis

ChIP-seq and ATAC-seq paired-end sequenced reads were trimmed using Trimmomatic/0.36 (31) and aligned to the mouse mm10 reference genome using Bowtie/2.3.5 (–very-sensitive –no-mixed –no-unal) (32). Bam files were created using view from SAMtools/1.9 (33) and duplicated reads were removed using MarkDuplicates from

Picard/2.17.3 (Broad Institute). Coverage tracks were created using bamCoverage (–normalizeUsing RPKM -bs 10 -e) from deepTools/3.3.0 (34). Peaks were called against sequenced input DNA or control Flag ChIP using callpeaks (-f BAMPE -p (1e-3) -g mm) from MACS/2.1.1 (35). Supplementary Table S2 lists the parameters of ChIPseq and ATACseq experiments. Categories of Pax7 binding sites are based on already published lists (4) but refined as described in Supplementary Table S3 to increase their homogeneity. Pax7 ChIP-seq peaks were categorized using the absence or presence, before or after Pax7 expression, of overlapping peaks from: H3K4me1 ChIP-seq to assess enhancer marking; p300 ChIP-seq to assess enhancer activation; and ATAC-seq to assess DNA accessibility. SMC1 ChIP-seq was used to avoid insulator contamination in the resistant category. In order to minimize threshold effects, the absence of a peak requires no peak to be detected with a low detection threshold whereas its presence requires its detection with a higher threshold.

Motifs analysis

De novo motif analysis of Pax7 ChIP-seq data was performed on 60 bp surrounding peak summits. Each subset of Pax7 peak were processed using HOMER *de novo* motif search (36) and MEME (http://meme-suite.org/tools/meme-chip), in order to find the motifs with the most significant *P*-value. Graphical representations of the position weight matrices obtained from these analyses were generated using a custom Python script for motif logos. Motifs frequencies were analyzed using a custom Python based program. Using uncovered matrices, motif scores for each Pax7 binding site were computed with a custom Python script in a 100 bp window around Pax7 peak summits.

Whole-genome bisulfite sequencing (WBGS)

The CpG methylation levels were represented using the published data from (4).

RESULTS

Strongest in vitro Pax7-binding motif is enriched at pioneered target sites

Prior studies (reviewed in (17)) on the DNA binding properties of the related Pax TFs Pax3 and Pax7 (Figure 1A) revealed cognate binding sites for either of their two DBDs, the PD or the HD domain (Figure 1B). Genome-wide assessment of these TF binding sites obtained by ChIPseq showed enrichment of the PD target sequence core motif TGAC and of different associations of the TAAT motif recognized by the HD (5,37). Prior work identified the H.I2 palindrome as a binding site for Pax3 and Pax7 dimers (38,39). The H.I-2 overlapping TAATTA motif was identified in ChIPseq data for Pax7 (5) but its significance remains poorly explored. We also reported that target enhancers pioneered by Pax7 show enrichment of the composite motif that contains juxtaposed PD and HD motifs (5). Extensive characterization of the Pax7 ChIPseq peaks defined different genomic recruitment sites (4). These included binding at putative enhancer sites that are already

accessible before Pax7 expression (Constitutive sites), at sites where Pax7 binding results in transcriptional activation revealed by p300 recruitment (activated sites/before Pax7: low H3K4me1 and ATAC, no p300/after Pax7: high H3K4me1 and ATAC, p300 present), at Resistant sites (before and after Pax7: no H3K4me1, ATAC or p300), and finally at Pioneered sites (before Pax7: no H3K4me1, ATAC or p300 / after Pax7: high H3K4me1 and ATAC, p300 present) or Primed sites (before Pax7: no H3K4me1, ATAC or p300/after Pax7: single peak low H3K4me1 and weak ATAC, no p300). Pax7 motif frequencies for these different subcategories show that the fully activated Pioneer sites have the highest frequency for all motifs (Figure 1C, D) and this correlates with this category of peaks exhibiting the strongest Pax7 ChIPseq signals of all categories (4) as supported by the highest median number of reads or *P* values for this subset (Figure 1C). The Pioneer subset also has the highest combined frequencies of all different Pax7 motifs (Figure 1E) consistent with the frequent occurrence of more than one motif under each peak (Figure 1F).

In order to compare Pax7 binding to each target type motifs, we did electrophoretic mobility shift assays (EMSA) using DNA probes corresponding to each motif (Figure 1B and Supplementary Figure S1A) and nuclear extracts from AtT20 cells expressing a Flag-tagged Pax7 or an empty vector (Neo). Complexes present in EMSA using Pax7 compared to Neo cells revealed a single complex present with all probes (arrowhead in Figure 1G). The apparent strength/stability of this Pax7 complex is however very different for different probes: the highest apparent affinity site is the composite HP.E0 site followed, in decreasing order, by the inverted TAATTA palindrome H.I-2, the PD motif, the TAAT palindrome H.I2, and the single HD. The only other Pax7-specific complex is observed with the H.I2 probe (asterisk in Figure 1G). This slower migrating band is a likely Pax7 dimer as this sequence was previously shown to form dimers with both Pax3 and Pax7 (38). We did competitive binding experiments in order to assess the relative affinities of complexes formed with the different probes and this supported the interpretation of the highest affinity for the composite followed by PD and HD sites (Supplementary Figure S1B and C). The presence of Pax7 in the purported Pax7-DNA complexes was ascertained by supershift EMSA experiments using the Flag antibody: both monomer and dimer Pax7:DNA complexes were indeed confirmed to contain the Flag-Pax7 chimeric protein (Supplementary Figure S1D).

Prevalent role of Pax7 paired (PD) domain for DNA binding

In order to assess the relative roles of PD and HD domains for DNA binding, we produced and purified polypeptides containing different Pax7 domains from over-expressing *E. coli*. Whereas the PD domain (residues 35–158 of mouse Pax7) is sufficient for in vitro DNA binding to the composite HP.E0 and PD probes, the HD domain (aa 215–273) does not bind significantly at the concentrations shown but did at 1000 times greater concentration (Figure 2A, lanes 1, 3, 5, 7 and not shown). Addition of the linker sequence present between PD and HD to the PD polypeptide appeared to stabilize PD probe binding and to result in a
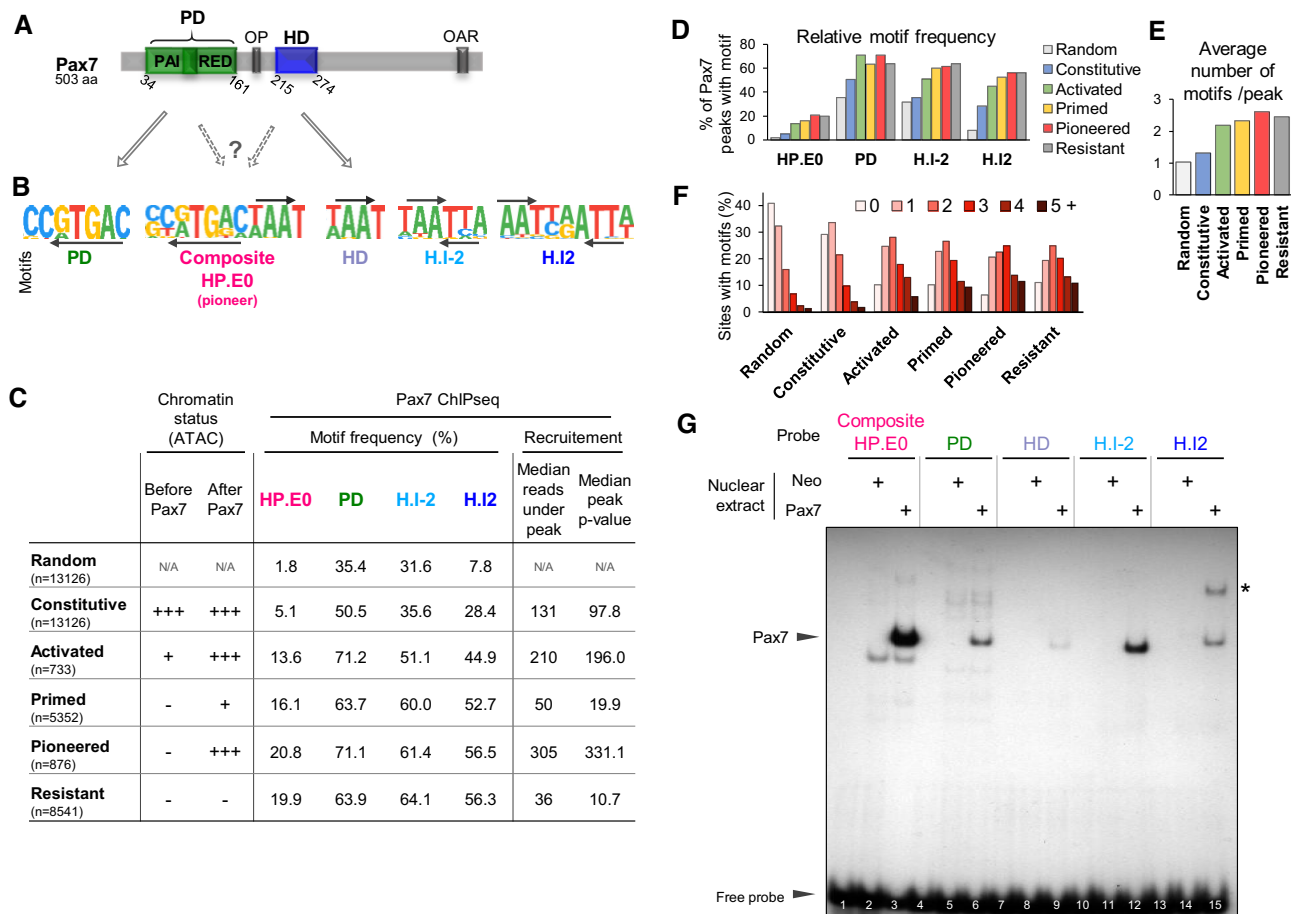
Figure 1. Contribution of paired (PD) and homeodomain (HD) DNA binding domains to Pax7 binding. (**A**) Schematic representation of mouse Pax7 structure (Uniprot P47239-1) including the PD and HD DNA binding domains. The positions of two other conserved domains is also indicated: OP, a conserved octapeptide present in the linker region between the two DNA binding domains and OAR, a conserved tetradecapeptide initially identified in the transcription factors otp, aristaless and rax. (**B**) Consensus recognition motifs identified in Pax7 ChIPseq (5) and related to recognition by either PD or HD as suggested by arrows linking panels A and B. The composite HP.E0 motif contains everted sequences of the PD and HD recognition motifs and this composite was identified as enriched under Pax7 peaks that trigger pioneer chromatin opening (5). (**C**) Distribution of Pax7 interaction motifs in subsets of Pax7 ChIPseq peaks defined by their changes (or not) in pituitary AtT20 cells before/after Pax7 expression. These categories were defined and discussed extensively in (4), and the subsets presented here (n indicated between parentheses) were trimmed to only include peaks with stringently defined properties compared to other subsets. For reference, the chromatin accessibility (ATAC) status of these subsets is indicated. The average Pax7 recruitment strength for each subset is indicated by the median number of reads and *P* values of peaks present in each subset. (**D**) Relative frequencies (% of peaks containing indicated motif) for each subset of peaks. (**E**) Average total number of all motifs per peak. (**F**) Distribution of number of motifs per peak for each subset of Pax7 ChIPseq peaks. (**G**) Binding of Pax7 to its different DNA targets shown in A above. Electrophoretic Mobility-Shift Assay (EMSA) was used to assess the in vitro binding of Pax7 stably expressed in AtT20 cells (4). Similar DNA probes (probe sequences in Supplementary Figure S1A) were incubated with either control AtT20-neo or AtT20-Pax7 cell nuclear extracts for analyses by EMSA. A single specific band is observed (arrowhead) for each probe but with widely varying intensities. Only the H.I2 palindromic probe shows an additional band (*) that is likely constituted of Pax7 homodimers.

more compact faster migrating complex (lane 6 compared to 5) but the strongest binding is observed for a polypeptide containing both PD and HD (PD.HD, aa 35–273) together with intervening linker sequences (Figure 2A, lanes 4 and 8). On its own, the HD polypeptide exhibited binding only to the palindromic H.I2 probe compared to the core HD or to the HD H.I-2 probes (Figure 2A, lanes 19 compared to 11, 15). The PD:HD polypeptide showed the strongest binding to all probes although this is marginal for the simple HD motif (lanes 4, 8, 12, 16, 20). Further, this PD:HD polypeptide presents an additional band with the H.I2 palindromic probe (lane 20) consistent with homodimer binding reported for both Pax3 and Pax7 with this sequence (38). Collectively, these data indicate that the PD

domain is primarily responsible for high affinity DNA binding and that the HD domain may contribute stabilization and specificity of binding, in particular with the composite HP.E0 probe. On the other hand, the HD may increase the repertoire of Pax7 binding sites to include inverted palindromic target sequences of the H.I2 type.

Pax7 binds the composite site as monomer

Binding of Pax7 to the composite element may involve the two DBDs of the same Pax7 molecule or the PD of one associated as dimer with the HD of another Pax7 molecule. In order to address this question, we performed EMSA using AtT20-Pax7 nuclear extracts with and without addition of
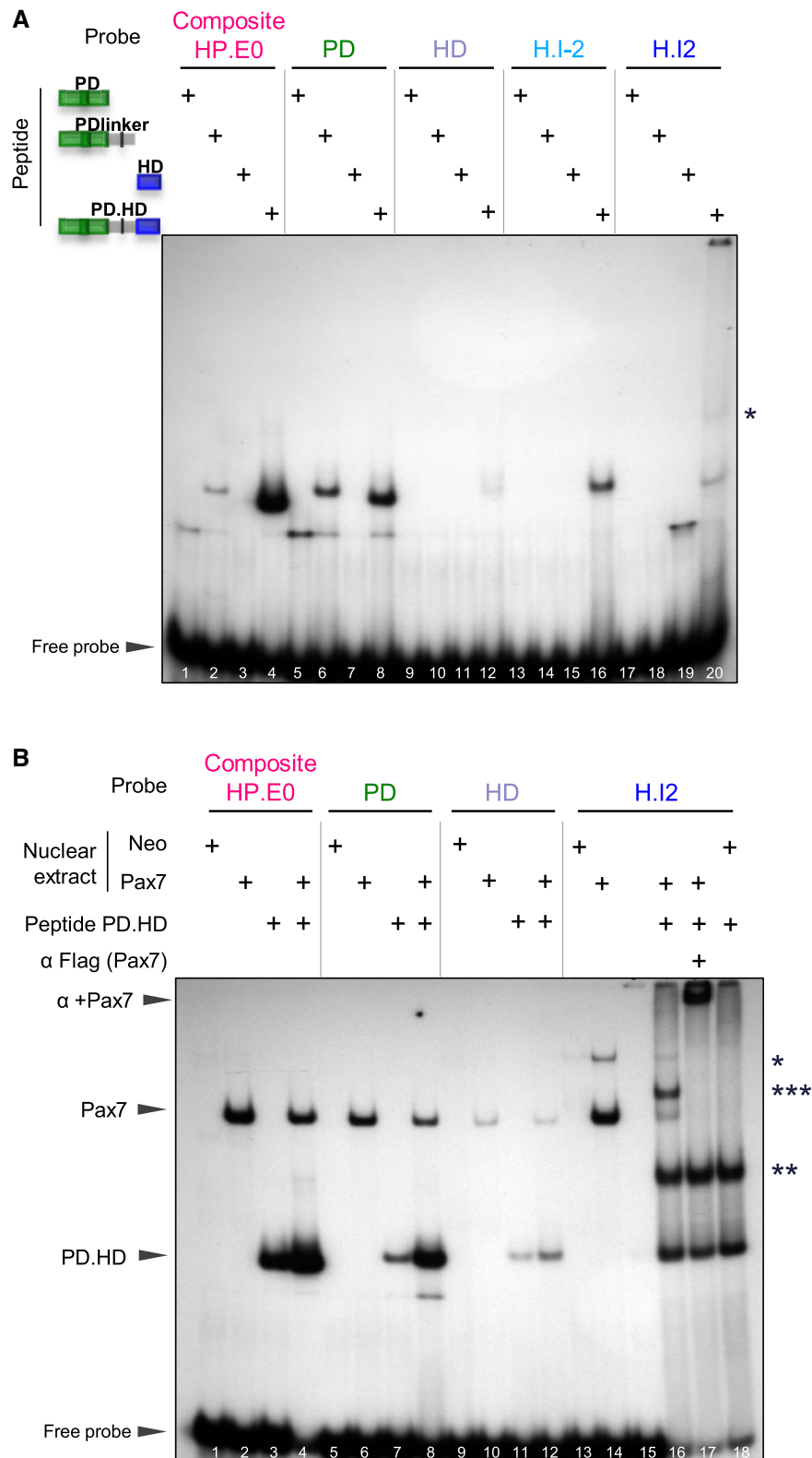
Figure 2. Relative roles of Pax7 PD and HD domains in DNA binding. (**A**) *In vitro* binding (EMSA) of various Pax7 polypeptides produced in *E. coli* to contain the represented portions of the Pax7 DNA binding domains. The purified polypeptides were incubated with ^{32}P-labeled probes before electrophoresis. For each probe, the strongest band likely represents monomer Pax7 polypeptide binding except for the H.I2 probe where a likely dimer band is observed (*). (**B**) EMSA mixing experiment showing that Pax7 monomers bind the composite site. EMSA was performed for each DNA probe using either AtT20 nuclear extracts expressing (or not, neo) Pax7 with/without the PD.HD polypeptide as indicated. Arrowheads indicate the positions of wild-type Pax7 and PD.HD polypeptide monomer complexes. For the H.I2 probe, dimer bands are also observed for Pax7 (*) and PD.HD (**) together with appearance of a new band (***) in the combined reaction (lane 16) that likely represents a heterodimer between intact Pax7 and PD.HD polypeptide. The presence of Flag-Pax7 in complexes is ascertained by supershifts with the Flag M2 antibody.

the bacterially-produced PD.HD polypeptide. These mixing experiments (Figure 2B) revealed independent binding of the PD.HD polypeptide and AtT20-expressed Flag-Pax7 to the composite (lanes 2–4), PD (lanes 6–7) and HD (lanes 11–13) probes. In contrast, binding of the full-length Pax7 to the palindromic H.I2 probe revealed two bands deemed to represent monomers and dimers (Figure 2B, lane 14). The PD.HD polypeptide also formed two complexes with the slower migrating band (labelled **) deemed to represent a dimeric complex that is not observed with the other probes (Figure 2B, lane 18). When both PD.HD polypeptide and full-length Pax7 are present together, a new band (labelled ***) was observed and interpreted as consisting of one molecule of full-length Pax7 together with one molecule of the PD.HD (Figure 2B, lane 16). This new complex together with the Pax7-containing complexes were all supershifted by addition of the anti-Flag antibody confirming presence of the AtT20-expressed Pax7 in these complexes (Figure 2B, lane 17). Taken together, these experiments are consistent with the interpretation that a monomer of Pax7 is sufficient to bind the composite target sequence with the PD and HD of the same Pax7 molecule likely to participate in close interactions in order to bind their juxtaposed target DNA sequences.

Cytosine methylation does not interfere with Pax7 DNA binding

Prior to Pax7 binding, the pioneered enhancers are within nucleosomal heterochromatin (4) as determined by MnaseSeq (Supplementary Figure S2). Upon Pax7 binding, nucleosomes are displaced from the Pax7 pioneered sites but not from other subsets of Pax7 binding sites (Supplementary Figure S2), and this correlates with the loss of histone H3 (4). The Pax7 pioneered sites are also within heterochromatin with high levels of CpG DNA methylation and this includes CpG residues that are within composite sites in some cases (4). These enhancers acquire stable epigenetic memory through demethylation of cytosine residues within targeted enhancers (Figure 3A, B). Indeed, the average level of CpG methylation drops from ∼80% to ∼30% after Pax7-dependent chromatin opening at the pioneered enhancers (Figure 3B). These data suggest that cytosine methylation of DNA target sequences does not interfere with Pax7 interaction, and we directly assessed this by producing DNA probes with cytosine methylation of the CpG dinucleotide present within the composite Pax7 binding site (Figure 3C). Probes containing methylcytosine on the sense (S, lane 6), the antisense (AS, lane 9) or both strands (lane 12) were compared to an unmethylated probe for Pax7 binding and Pax7 was found to bind equally well to all probes (Figure 3D). This was also found to be the case for Pax7 binding to the methylated PD probe which showed weaker, but nonetheless methylation-independent, binding (Figure 3D, lanes 13–25).

The ability of Pax7 to bind its composite and PD sequences independently of cytosine methylation status is entirely consistent with the idea that this pioneer factor can recognize its target sequence within 'closed' nucleosomal heterochromatin that is heavily methylated on cytosines.

Both Pax7 DNA binding domains are required for pioneer activity

To ascertain the role of the two Pax7 DBDs in transcription and pioneer action, we introduced single amino acid mutations in either PD or HD (Figure 4A). For the PD domain, the amino acid substitution R56L in the α1 helix of the PAI domain was derived from a Waardenburg syndrome mutation identified in the related human PAX3 gene (40,41), and shown to be a loss-of-function mutation (42,43). Another PD mutation G43A was introduced upstream of the α helices of the PAI domain; this mutation was previously shown to loose PD, but not palindromic HD, sequence binding (19). For the HD domain, the critical recognition residue Serine 50 of the HD was changed to glutamic acid (S264E). For comparisons, Supplementary Figure S4 provides an alignment of sequences for mouse Pax7 (numbering used in the present work), human PAX3 and human PAX6. These mutations were introduced in Flag-Pax7 and stably expressed in AtT20 cells as previously for Pax7 (5). The three Flag-Pax7 mutant proteins are expressed at similar levels (Supplementary Figure S3A). *In vitro* (EMSA), the R56L and G43A PD mutants did not bind the PD probe significantly (Supplementary Figure S3B, lanes 8 and 9). Strikingly, all three mutations had much less effect on binding to the composite probe (Supplementary Figure S3B, lanes 2–5). For HD probes H.I2 and H.I-2, the HD S264E mutant lost most binding activity but so did the PD mutants (Supplementary Figure S3B, lanes 13–15 and 18–20).

Since the crystal structure of hPAX6 showed that the Arg residue corresponding to R56L (labeled R23 in Xu et al. (44)) contacts the DNA backbone (but not the methylation site per se) at the level of the CpG dinucleotide present within the PD motif, we assess the hypothesis that the mutations may selectively affect binding of the methylated target sequences. We thus tested all three mutants for binding of the methylated (on both strands) composite and PD sites compared to unmethylated DNA (Figure 4B); other CpG present in our probes are outside of conserved DNA sequences identified at Pax7 ChIPseq peaks. DNA methylation of the CpG dinucleotide present within the PD sequence of either probe did not affect binding by any mutant.

The transcriptional potential of these mutants was assessed by co-transfection into AtT-20 cells using a Luciferase reporter driven by six copies of the composite element or by the PC2 gene 1.4 kb enhancer located at −146 kb upstream of TSS (5). For the composite element reporter, transcriptional activity was affected by the R56L mutation but not by the G43A or S264E mutations. Except for R56L, these results (Supplementary Figure S3C) are consistent with the in vitro binding results using the composite probe (Supplementary Figure S3B). In contrast, the activity of the PC2 enhancer reporter was dependent on both PD and HD as evidenced by significant activity losses for all three mutants (Supplementary Figure S3C). The dependence on both the PD and HD for activity at this natural enhancer is thus more complex than the simpler *in vitro* binding patterns observed in EMSA.

The impact of the DBD mutations on *in vivo* genomic recruitment was assessed by ChIPseq (Figure 4C). It is quite striking to observe that key amino acid substitutions in ei-
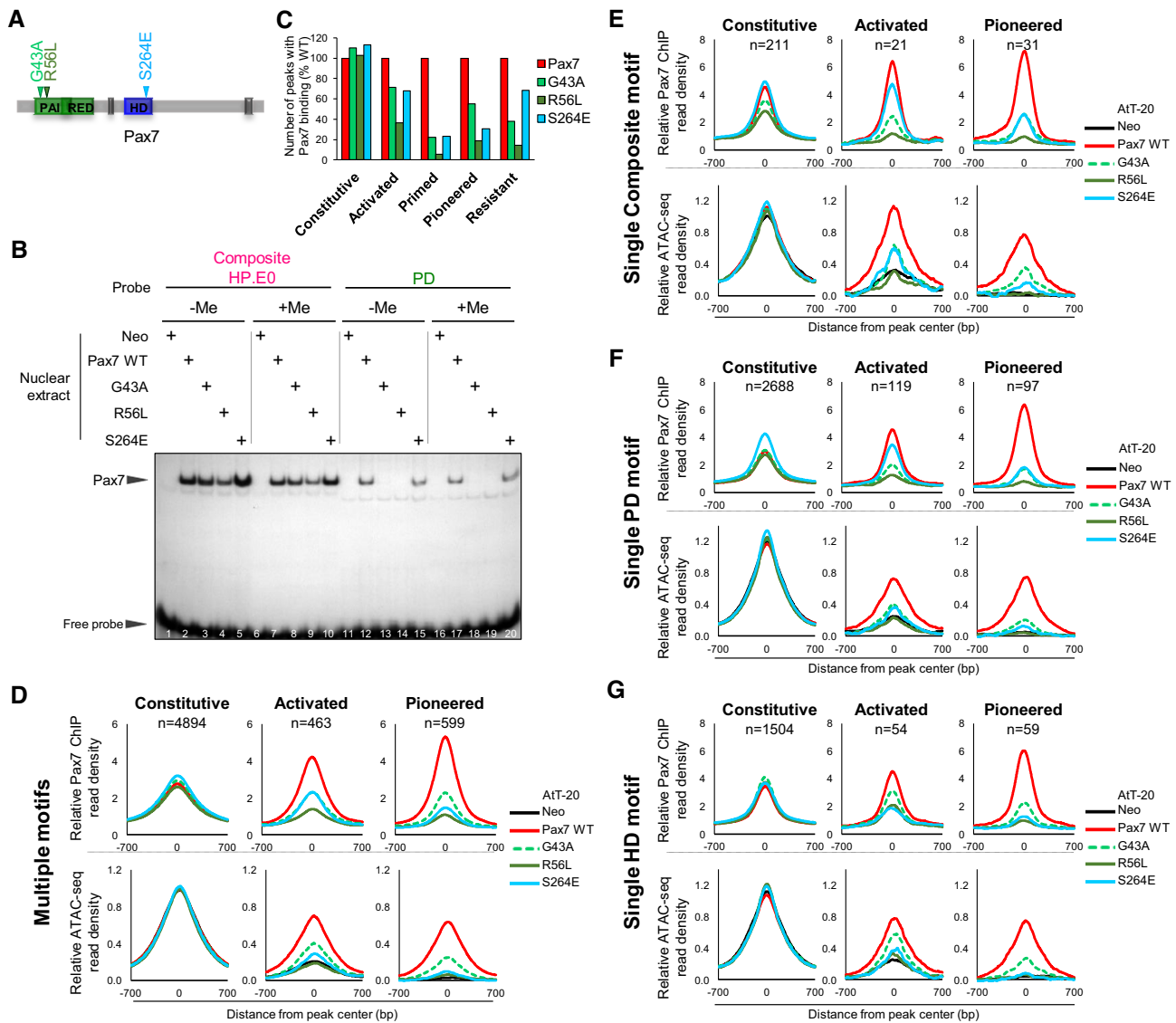
Figure 3. Pax7 binding is independent of DNA methylation. (**A, B**) Average profiles of DNA methylation (% methylated CpG) around Pax7 ChIPseq peaks present at Constitutive (A) and Pioneered (B) enhancers. Data are shown for control AtT20-neo compared to AtT20-Pax7 cells and aligned with the average Pax7 ChIPseq profiles. (**A**) DNA sequences of probes used in EMSA to test the effect of DNA methylation on Pax7 binding. Methylated cytosine (C) residues are shown on sense (S) strand. (**B**) *In vitro* binding (EMSA) of Pax7 to either composite or PD probes methylated or not on sense (S), antisense (AS) or both (Me) strands.

Figure 4. Intact and functional PD and HD domains are both required for recruitment and chromatin opening at Pax7 pioneer sites. (**A**) Schematic representation of Pax7 showing three amino acid substitutions assessed for recruitment by ChIPseq and chromatin opening by ATACseq. (**B**) Effect of three Pax7 mutations on ChIPseq recruitement at different subset of peaks as indicated. The number of peaks (% compared to wild-type Pax7) that retained significant ($P < 10^{-5}$) binding are shown. (**C**) *In vitro* binding (EMSA) of Pax7 and its mutants to either Composite or PD probes methylated or not (+/−Me) on both strands. (**D**) Average profiles of Pax7 ChIPseq and ATACseq signals for the indicated subsets of Pax7 peaks and for the subsets of peaks in each category that have multiple Pax7 recognition motifs. (**E**) Average profiles for the subsets of Pax7 peaks of the indicated categories that only contain one composite motif. (**F**) Average profiles for the subsets of Pax7 peaks of the indicated categories that only contain one PD motif. (**G**) Average profiles for the subsets of Pax7 peaks of the indicated categories that only contain one HD motif.

ther PD or HD failed to affect the recruitment of Pax7 to the Constitutive enhancers that are already in an open chromatin state: this likely reflects the strong contribution of protein-protein interactions for recruitment of Pax7 in the context of active enhancers. In contrast, recruitment to the Pioneered enhancers is severely curtailed by all three mutations (loss of recruitment at ∼80% of sites for R56L, at ∼70% sites for S264E and ∼50% sites for G43A) with the G43A PD substitution having a partial effect compared to the R56L substitution (Figure 4C and Supplementary Figure S3D). Thus, both DBDs must be intact for pioneer site binding whether considering the fully activated Pioneer sites or the Primed sites. Intermediate effects of the different mu-

tations are observed at the group of so-called Activated enhancers, a group of enhancers with open chromatin where the co-activator p300 is recruited after Pax7 binding. Collectively, these data indicate that the G43A mutant is a partial loss-of-function whereas the other mutants appear complete nulls. Finally, it is noteworthy that binding of the Pax7 mutants to the group of Resistant sites, binding sites that show no chromatin change upon Pax7 binding, is markedly impaired by the PD mutations but relatively unaffected by the HD S246E mutation. This data indicates that the PD domain is primarily involved in scanning heterochromatin sites for binding and that this function is relatively independent of the HD.

The loss of recruitment observed in ChIPseq for the mutants (assessed as the presence of peaks above threshold) is significant for the Pioneered, Primed and Activated sites but it is not complete (Figure 4C). In order to ascertain the meaning of this residual recruitment, we selected subsets of enhancers (Constitutive, Pioneered or Activated, Figure 4E–G) based on presence of a single Pax7 recognition motif, either composite, PD or HD motif. For comparisons, average profiles of Pax7 ChIPseq peaks containing multiple motifs are shown in Figure 4D. The Constitutive Pax7 peaks are relatively insensitive to either PD or HD domain mutations irrespective of the presence of multiple motifs (Figure 4D) or of only one motif (Figure 4E–G). In order to correlate Pax7 recruitment with pioneering activity, we performed ATACseq for the Pax7 DBD mutants and compared them to the ATAC profiles of cells expressing wild-type Pax7. These profiles (Figure 4D–G, bottom panels) illustrate the open chromatin status of Constitutive enhancers independently of Pax7 recruitment.

In sharp contrast, the Pioneered peaks show dependence on intact DBDs for both PD and HD, as much for Pax7 recruitment as for chromatin opening (ATAC).

Recruitment to the Activated subset is more dependent on DNA recognition specificity. Indeed, peaks that have either single composite (Figure 4E) or PD motifs (Figure 4F) show relatively unaltered Pax7 recruitment for the HD mutant S264E but this is only marginally reflected in ATACseq signals which are similarly decreased by both PD and HD mutations. For sites that have a single HD motif, both Pax7 recruitment and ATAC signals are affected by mutations in either PD or HD. Specific loci examples of these effects are provided in Supplementary Figure S5. Collectively, these results indicate that PD and HD domains are required to stably bind Pioneer sites and trigger chromatin opening. For Activated sites that already exhibit some chromatin opening (ATAC signal), there is a similar dependence on both domains for binding targets with a single HD motif whereas single PD or composite motif-containing targets only show dependence on the intact PD but not intact HD.

We then asked whether gene activation shows similar dependence on intact PD and HD domains as revealed in ChIP recruitment. This was assessed by RT-qPCR measurements of mRNA levels for gene targets that are either dependent on regular transcriptional activation or on pioneering mechanism (Figure 5A and Supplementary Figure S6A). We previously showed rapid transcriptional response to Pax7 of genes that are subject to transcriptional activation compared to the slower (days) response of genes that require pioneer-dependent enhancer opening (4). The mRNA accumulation time courses of genes assessed for the Pax7 mutants are illustrated in Supplementary Figure S5A. Consistent with the Pax7 recruitment and ATAC data, activation/transcription of genes dependent on pioneer action show dependence on both intact PD and intact HD (Figure 5A). Activation of gene transcription for gene targets of the Activated class of enhancers is more heterogeneous with some showing dependence on both DBDs (*Pcdh9*, *Pde2a*, *Gas6*) and some with lesser dependence on the HD domain (*Gprc5*, *Pgc*). Whereas virtually all Pax7 target genes are regulated by Pax7-dependent enhancers, one notable exception is the *Oacyl* gene where Pax7 exerts

its pioneer action at the promoter (Supplementary Figure S6B); this gene shows a quicker response to Pax7 expression compared to most other Pax7 pioneered target genes (Supplementary Figure S6A).

Collectively, the mutant analyses reveal a strong dependence on both the PD and HD for pioneer activity independently of underlying DNA motifs present in pioneered enhancers. This may relate to a unique arrangement of the PD and HD consistent with interaction at composite binding sites.

DISCUSSION

This work has queried the requirements for selection of DNA target sites susceptible to pioneer factor chromatin opening. By comparing the DNA sequence motifs present under pioneered sites with those of other subsets of Pax7 binding sites together with mutations within each Pax7 DBD, the unique requirements for pioneer action suggest that structural features required for pioneer activity are far more stringent than for recruitment at enhancers within pre-opened chromatin. Much more than simple DNA sequence recognition is involved for recruitment of TFs and pioneers, whether considering recruitment of the former at different enhancer states or, of the latter, in different chromatin environments.

What may distinguish the Pax7 pioneered sites compared to other binding/recruitment sites? Compared to all other Pax7 binding sites, the pioneered subset is enriched in high affinity binding sites such as the composite motif, but this site is only present in about 20% of all pioneered enhancers. The pioneered sites are also enriched for the PD motif, but this enrichment is not strikingly different when compared to its abundance in other subsets of Pax7 peaks. Nonetheless, Pax7 binding is clearly much stronger at the pioneer subset compared to other subsets (4). Thus, it may be the number of motifs that results in stronger recruitment, for example when comparing the pioneered subset with the constitutive enhancers, but it is not the specific DNA sequence of the motifs present in different subsets that account for their different properties. This argument is strongly supported by the comparison of specific motifs and their abundance in the subset of Resistant compared to Pioneered sites: both subsets have very similar distribution of all PD and HD binding motifs. This observation clearly points towards heterochromatin organization as the defining property that establishes susceptible and resistant heterochromatin for pioneering.

A unique property of pioneer factors appears to be their ability to gain access to their target DNA sequence within nucleosomal and compacted heterochromatin. The molecular basis for this ability remains unexplained although some hypotheses were put forward. For example, it was proposed that FoxA pioneers may have this ability through molecular mimicry where a portion of the winged helix domain may mimic a structure within linker histone H1 (7). However, this model is not likely to apply to many other pioneers. The FoxA pioneers also have the ability to bind low-affinity sites within heterochromatin and the ability to recognize these low sequence conservation sites was proposed to represent a mechanism to scan putative sites within heterochromatin (21). Further, the presence of a core histone interaction do-
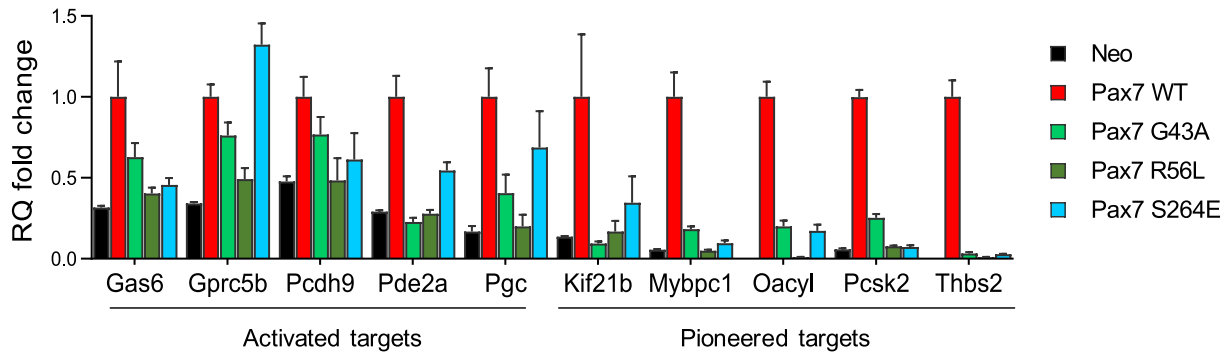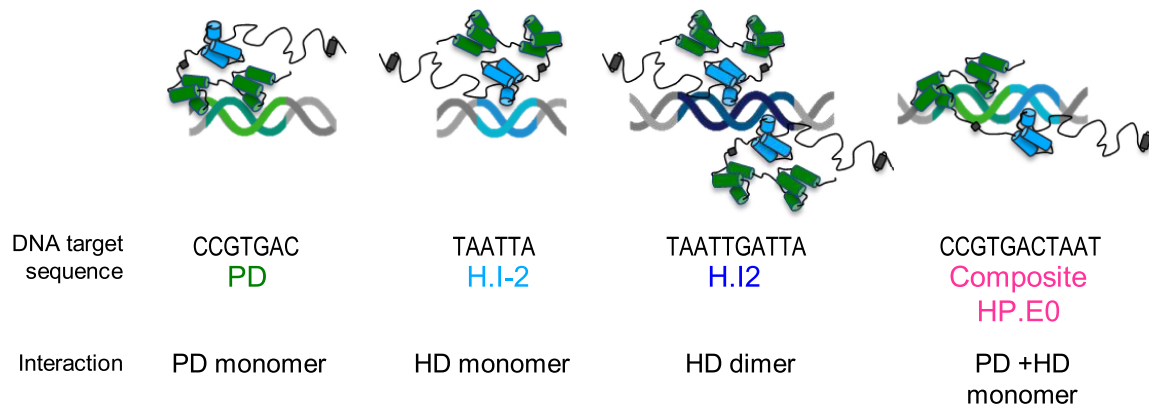
Figure 5. Mutations of either the PD or HD prevent gene activation at targets that require pioneer action whereas they have variable effects on transcriptional target genes. (**A**) RT-qPCR measurements of mRNA levels for genes subject to activation through either Pax7 pioneering or regular transcriptional activation as indicated. The gene activation time courses (slow or rapid, respectively) for these genes is shown in Supplementary Figure S6A. Data are presented for wild-type Pax7 in comparison with mutations in either PD or HD DNA binding domains. (**B**) Models of Pax7 interaction with its different DNA target sequences.

main within FoxA factors (23) clearly supports the hypothesis of direct chromatin interaction as a unique feature of pioneers providing them access to heterochromatin. Be that as it may, this fails to adequately explain how specific sites are selected for pioneering. It is quite striking that the ability for Pax7 recruitment to the weak binding sites of the Resistant subset is almost entirely dependent on the PD, but not the HD (Figure 4C). This preference is not reflected in an enrichment for the PD motif within the Resistant subset compared to HD motifs and thus likely reflects a property of the Pax7 PD to recognize heterochromatin.

Whereas the PD appears predominantly responsible for scanning heterochromatin sites, chromatin opening and gene activation by Pax7 pioneering is completely dependent on both intact PD and HD. This dependence is clear for both the fully activated Pioneered sites as well as for the primed sites: hence, this dependence probably reflects requirements for the initial step of the chromatin opening mechanism. The dependence on both intact DBDs for pioneering is not strictly correlated with the presence of cognate DNA sequence motifs for these domains within the pi-

oneered subset of enhancers, although this subset has the highest enrichment for all motifs and particularly the high-affinity composite (HP.E0) sites. Given the higher affinity of the composite site for Pax7 binding in vitro and its association with a Pax7 monomer (Figure 5B), but recognizing that many pioneered enhancers do not contain this high-affinity site, it is reasonable to propose that a unique conformation of the two DBDs relative to each other is associated with pioneering ability, and that this conformation can be achieved through different combinations of DNA sequence motif arrangements. This hypothetical conformation illustrated over the composite motif may differ compared to the Pax7 dimers interacting with the palindromic H.I2 target or even compared to simple PD or HD motifs (Figure 5B).

The situation at pioneered enhancers is in striking contrast with the absence of dependence on either intact PD or HD for recruitment to the Constitutive subset of enhancers that were already in an open and active chromatin state before Pax7 expression. This independence towards intact DBDs is correlated with the lesser abundance of PD and HD motifs within the Constitutive subset of enhancers

and likely reflects the greater importance of protein-protein interactions at these enhancers that are already occupied by an array of different transcription factors. It is noteworthy that the subset of Pax7 targets that exhibits the greatest dependence on cognate DNA sequence motifs are found within the Activated subgroup of enhancers, those enhancers that are in a primed state at the onset but become fully activated upon Pax7 binding as revealed by their recruitment of the co-activator p300. For this subset, the presence of a cognate DNA motif is most of the time correlated with the loss of this binding for the Pax7 mutant that contains a substitution in the corresponding DBD. It is noteworthy that target DNA sequence recognition is most important for this group of enhancers that are awaiting for a cognate TF for full activation. In this instance, the pioneer properties are likely not required as the chromatin is already slightly accessible, ie in the primed state.

In summary, the present work defined unique contributions of the two Pax7 DBDs towards pioneer action. First, pioneer factors demarcate themselves compared to other TFs by the ability to recognize targets within heterochromatin and this property appears to be associated with the PD. Second, pioneered sites exhibit the strongest recruitment in ChIPseq and this correlates with higher frequency of cognate DNA motifs and of sites of high affinity. Finally, a unique conformation that requires both the PD and the HD of Pax7 is required for chromatin opening at pioneer sites and gene activation.

DATA AVAILABILITY

Sequencing data are available under GSE87185 and GSE173743.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Mayran,A. and Drouin,J. (2018) Pioneer transcription factors shape the epîgenetic landscape. *J. Biol. Chem.*, **293**, 13795–13804.
2. Klemm,S.L., Shipony,Z. and Greenleaf,W.J. (2019) Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.*, **20**, 207–220.
3. Zaret,K.S. and Mango,S.E. (2016) Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr. Opin. Genet. Dev.*, **37**, 76–81.
4. Mayran,A., Khetchoumian,K., Hariri,F., Pastinen,T., Gauthier,Y., Balsalobre,A. and Drouin,J. (2018) Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. *Nat. Genet.*, **50**, 259–269
5. Budry,L., Balsalobre,A., Gauthier,Y., Khetchoumian,K., L'Honore,A., Vallette,S., Brue,T., Figarella-Branger,D., Meij,B. and Drouin,J. (2012) The selector gene Pax7 dictates alternate pituitary cell fates through its pioneer action on chromatin remodeling. *Genes Dev.*, **26**, 2299–2310.
6. Mayran,A., Sochodolsky,K., Khetchoumian,K., Harris,J., Gauthier,Y., Bemmo,A., Balsalobre,A. and Drouin,J. (2019) Pioneer and nonpioneer factor cooperation drives lineage specific chromatin opening. *Nat. Commun.*, **10**, 3807.
7. Cirillo,L.A., McPherson,C.E., Bossard,P., Stevens,K., Cherian,S., Shim,E.Y., Clark,K.L., Burley,S.K. and Zaret,K.S. (1998) Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J.*, **17**, 244–254.
8. Fernandez Garcia,M., Moore,C.D., Schulz,K.N., Alberto,O., Donague,G., Harrison,M.M., Zhu,H. and Zaret,K.S. (2019) Structural features of transcription factors associating with nucleosome binding. *Mol. Cell*, **75**, 921–932.
9. Cirillo,L.A., Lin,F.R., Cuesta,I., Friedman,D., Jarnik,M. and Zaret,K.S. (2002) Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell*, **9**, 279–289.
10. Iwafuchi-Doi,M., Donahue,G., Kakumanu,A., Watts,J.A., Mahony,S., Pugh,B.F., Lee,D., Kaestner,K.H. and Zaret,K.S. (2016) The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol. Cell*, **62**, 79–91.
11. van Oevelen,C., Collombet,S., Vicent,G., Hoogenkamp,M., Lepoivre,C., Badeaux,A., Bussmann,L., Sardina,J.L., Thieffry,D., Beato,M. *et al.* (2015) C/EBPalpha activates pre-existing and de novo macrophage enhancers during induced pre-B cell transdifferentiation and myelopoiesis. *Stem Cell Rep.*, **5**, 232–247.
12. Boller,S., Ramamoorthy,S., Akbas,D., Nechanitzky,R., Burger,L., Murr,R., Schubeler,D. and Grosschedl,R. (2016) Pioneering activity of the C-terminal domain of EBF1 shapes the chromatin landscape for B cell programming. *Immunity*, **44**, 527–541.
13. Li,R., Cauchy,P., Ramamoorthy,S., Boller,S., Chavez,L. and Grosschedl,R. (2018) Dynamic EBF1 occupancy directs sequential epigenetic and transcriptional events in B-cell programming. *Genes Dev.*, **32**, 96–111.
14. Wapinski,O.L., Lee,Q.Y., Chen,A.C., Li,R., Corces,M.R., Ang,C.E., Treutlein,B., Xiang,C., Baubet,V., Suchy,F.P. *et al.* (2017) Rapid chromatin switch in the direct reprogramming of fibroblasts to neurons. *Cell Rep.*, **20**, 3236–3247.
15. Pataskar,A., Jung,J., Smialowski,P., Noack,F., Calegari,F., Straub,T. and Tiwari,V.K. (2016) NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. *EMBO J.*, **35**, 24–45.
16. Takaku,M., Grimm,S.A., Shimbo,T., Perera,L., Menafra,R., Stunnenberg,H.G., Archer,T.K., Machida,S., Kurumizaka,H. and Wade,P.A. (2016) GATA3-dependent cellular reprogramming requires activation-domain dependent recruitment of a chromatin remodeler. *Genome Biol.*, **17**, 36.
17. Mayran,A., Pelletier,A. and Drouin,J. (2015) Pax factors in transcription and epigenetic remodelling. *Semin. Cell Dev. Biol.*, **44**, 135–144.
18. Jun,S. and Desplan,C. (1996) Cooperative interactions between paired domain and homeodomain. *Development*, **122**, 2639–2650.
19. Underhill,D.A. and Gros,P. (1997) The paired-domain regulates DNA binding by the homeodomain within the intact Pax-3 protein. *J. Biol. Chem.*, **272**, 14175–14182.
20. Apuzzo,S., Abdelhakim,A., Fortin,A.S. and Gros,P. (2004) Cross-talk between the paired domain and the homeodomain of Pax3: DNA binding by each domain causes a structural change in the other domain, supporting interdependence for DNA Binding. *J. Biol. Chem.*, **279**, 33601–33612.
21. Soufi,A., Garcia,M.F., Jaroszewicz,A., Osman,N., Pellegrini,M. and Zaret,K.S. (2015) Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell*, **161**, 555–568.
22. Donaghey,J., Thakurela,S., Charlton,J., Chen,J.S., Smith,Z.D., Gu,H., Pop,R., Clement,K., Stamenova,E.K., Karnik,R. *et al.* (2018)

Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat. Genet.*, **50**, 250–258.

23. Iwafuchi,M., Cuesta,I., Donahue,G., Takenaka,N., Osipovich,A.B., Magnuson,M.A., Roder,H., Seeholzer,S.H., Santisteban,P. and Zaret,K.S. (2020) Gene network transitions in embryos depend upon interactions between a pioneer transcription factor and core histones. *Nat. Genet.*, **52**, 418–427.

24. Rambaud,J., Desroches,J., Balsalobre,A. and Drouin,J. (2009) TIF1beta/KAP-1 is a coactivator of the orphan nuclear receptor NGFI-B/Nur77. *J. Biol. Chem*, **284**, 14147–14156.

25. Khetchoumian,K., Balsalobre,A., Mayran,A., Christian,H., Chénard,V., St-Pierre,J. and Drouin,J. (2019) Pituitary cell translation and secretory capacities are enhanced cell autonomously by the transcription factor Creb3l2. *Nat. Commun.*, **10**, 3960.

26. Vallette-Kasic,S., Couture,C., Balsalobre,A., Gauthier,Y., Metherell,L.A., Dattani,M. and Drouin,J. (2007) The TPIT gene mutation M86R associated with isolated ACTH deficiency interferes in protein:protein interactions. *J. Clin. Endocrinol. Metab*, **92**, 3991–3999.

27. Langlais,D., Couture,C. and Drouin,J. (2012) The Stat3/GR interaction code: predictive value of direct/indirect DNA recruitment for transcription outcome. *Mol. Cell*, **47**, 38–49.

28. Lerdrup,M., Johansen,J.V., Agrawal-Singh,S. and Hansen,K. (2016) An interactive environment for agile analysis and visualization of ChIP-sequencing data. *Nat. Struct. Mol. Biol.*, **23**, 349–357.

29. Thorvaldsdottir,H., Robinson,J.T. and Mesirov,J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform*, **14**, 178–192.

30. Cui,K. and Zhao,K. (2012) Genome-wide approaches to determining nucleosome occupancy in metazoans using MNase-Seq. *Methods Mol. Biol.*, **833**, 413–419.

31. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

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

33. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

34. Ramírez,F., Ryan,D.P., Grüning,B., Bhardwaj,V., Kilpert,F., Richter,A.S., Heyne,S., Dündar,F. and Manke,T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.*, **44**, W160–W165.

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

36. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.

37. Soleimani,V.D., Punch,V.G., Kawabe,Y., Jones,A.E., Palidwor,G.A., Porter,C.J., Cross,J.W., Carvajal,J.J., Kockx,C.E., van Ijcken,W.F. *et al.* (2012) Transcriptional dominance of pax7 in adult myogenesis is due to high-affinity recognition of homeodomain motifs. *Dev. Cell*, **22**, 1208–1220.

38. Wilson,D., Sheng,G., Lecuit,T., Dostatni,N. and Desplan,C. (1993) Cooperative dimerization of paired class homeo domains on DNA. *Genes Dev.*, **7**, 2120–2134.

39. Birrane,G., Soni,A. and Ladias,J.A. (2009) Structural basis for DNA recognition by the human PAX3 homeodomain. *Biochemistry*, **48**, 1148–1155.

40. Chalepakis,G., Goulding,M., Read,A., Strachan,T. and Gruss,P. (1994) Molecular basis of splotch and Waardenburg Pax-3 mutations. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 3685–3689.

41. Hoth,C.F., Milunsky,A., Lipsky,N., Sheffer,R., Clarren,S.K. and Baldwin,C.T. (1993) Mutations in the paired domain of the human PAX3 gene cause Klein-Waardenburg syndrome (WS-III) as well as Waardenburg syndrome type I (WS-I). *Am. J. Hum. Genet.*, **52**, 455–462.

42. Armoni,M., Quon,M.J., Maor,G., Avigad,S., Shapiro,D.N., Harel,C., Esposito,D., Goshen,Y., Yaniv,I. and Karnieli,E. (2002) PAX3/forkhead homolog in rhabdomyosarcoma oncoprotein activates glucose transporter 4 gene expression in vivo and in vitro. *J. Clin. Endocrinol. Metab.*, **87**, 5312–5324.

43. Fortin,A.S., Underhill,D.A. and Gros,P. (1997) Reciprocal effect of Waardenburg syndrome mutations on DNA binding by the Pax-3 paired domain and homeodomain. *Hum. Mol. Genet*, **6**, 1781–1790.

44. Xu,H.E., Rould,M.A., Xu,W., Epstein,J.A., Maas,R.L. and Pabo,C.O. (1999) Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev.*, **13**, 1263–1275.