Table 1 Clinical features of study population

	Type 1 autoimmune hepatitis	Drug-induced liver injury	*P*-value
Sample size, *n*	71	37	–
Age, years	58 (20–86)	53 (24–78)	0.029
Sex, female (%)	63 (89%)	23 (62%)	0.001
Revised scoring system proposed by the International Autoimmune Hepatitis Group			
Definite diagnosis	55 (77%)	–	–
Laboratory data			
WBC (/mm^3)	5300 (1600–9300)	5700 (1200–24500)	0.016
Hemoglobin (g/dL)	12.9 (8.7–16.1)	13.4 (9.0–16.6)	0.039
Platelet (×10^4/mm^3)	18.2 (7.4–46.0)	21.4 (4.4–32.0)	0.047
Bilirubin (mg/dL)	1.4 (0.5–24.9)	6.0 (0.3–26.7)	0.011
AST (IU/L)	288 (30–2466)	658 (99–13966)	0.003
ALT (IU/L)	337 (21–2377)	750 (54–4816)	<0.001
IgG (g/dL)	2.2 (1.0–4.9)	1.4 (0.8–2.4)	<0.001
Prothrombin activity (%)	78.8 (19.7–111.3)	75.4 (12.1–161.5)	0.25
ANA (%)			<0.001
<1:40	2 (3%)	25 (68%)	
≥1:40 and ≤1:80	28 (39%)	6 (16%)	
≥1:160	41 (58%)	6 (16%)	
Liver histology			
Acute hepatitis (%)	8 (11%)	–	–
Cirrhosis (%)	6 (8%)	–	–

ALT, alanine aminotransferase; ANA, antinuclear antibody; AST, aspartate aminotransferase; IgG, immunoglobulin G; WBC, white blood cell.

order to avoid inter-plate variability, we used a positive serum, assigned it 0.200 OD_{630nm}, and read the optical densities of all samples against this positive serum. Intra-assay variability was found to be 8.4%.

Previously, we showed that serum levels of anti-PD-1 antibodies in 62 healthy volunteers were a median of 0.033 (range, 0.002–0.144) OD_{630nm}.[3] According to our previous report,[3] the cut-off level in this study was represented by a mean absorbance +2 standard deviations in healthy volunteers (=0.086 OD_{630nm}).

Statistical analysis

Statistical analysis was performed using the SPSS statistical program (SPSS, Chicago, IL, USA).

Continuous variables were expressed as a median (range). Differences in continuous variables between two independent samples were evaluated by the Mann–Whitney *U*-test. Dichotomous variables were compared by the χ^2-test. Spearman's rank correlation coefficient was used to evaluate the consistency in the continuous variables between two independent samples. Univariate and multivariate Cox proportional hazard models were performed to identify factors associated with the later normalization of serum ALT levels. The variables, which showed $P < 0.1$ by univariate analysis, were included into the multivariate analysis. A cumulative incidence was analyzed using the Kaplan–Meier method, and the differences in the curves were evaluated using the log–rank test. The diagnostic accuracy of each factor was evaluated based on the area under the curve (AUC) using receiver–operator curve (ROC) analysis. The threshold of the reported *P*-values for significance was accepted as less than 0.05.

RESULTS

Anti-PD-1 antibody in type 1 AIH and DILI

SERUM LEVELS OF anti-PD-1 antibodies were significantly higher in type 1 AIH patients (0.071 [0.011–0.449] OD_{630nm}) than in DILI patients (0.023 [0.003–0.180] OD_{630nm}) ($P < 0.001$) (Fig. 1). When the cut-off level was represented according to the previous report (=0.086 OD_{630nm}),[3] positivity for serum anti-PD-1 antibody was shown in 38% (27/71) of type 1 AIH patients and 8% (3/37) of DILI patients ($P = 0.001$).

Two of three DILI patients positive for serum anti-PD-1 antibody had acute liver failure positive for serum ANA (Table 2).[10]

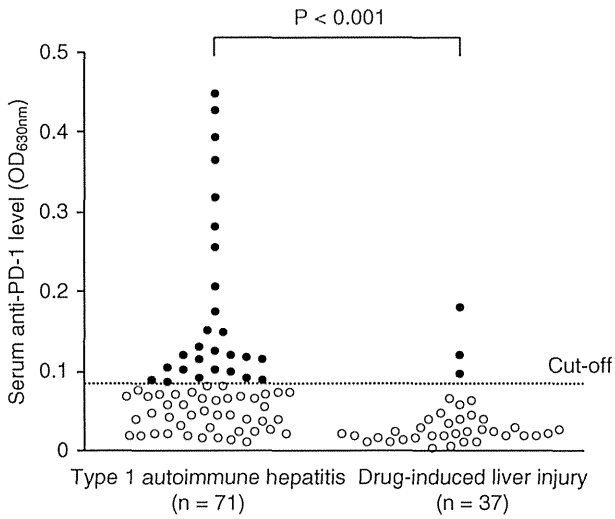

Figure 1 Serum levels of anti-programmed cell death-1 (anti-PD-1) antibodies in study populations. Closed circles show positivity for serum anti-PD-1 antibody. Open circles show negativity for serum anti-PD-1 antibody. According to the previous report,[3] the cut-off level was represented by a mean absorbance +2 standard deviations in healthy volunteers (=0.086 optical densities at 630 nm [OD_{630nm}]).

Anti-PD-1 antibody and clinical features of type 1 AIH

In type 1 AIH patients, serum levels of anti-PD-1 antibodies were inversely correlated with prothrombin activities ($\rho = -0.31$; $P = 0.008$), and prothrombin activities were lower in patients positive for serum anti-PD-1 antibody than in those negative for serum anti-PD-1 antibody (69.5% [19.7–102.6%] vs 86.1% [36.0–111.3%]; $P = 0.002$). Serum bilirubin levels tended to be correlated with serum levels of anti-PD-1 antibodies ($\rho = 0.23$; $P = 0.051$) and higher in patients positive for serum anti-PD-1 antibody (2.7 mg/dL [0.5–24.4] vs 1.2 mg/dL [0.5–24.9]; $P = 0.078$). But, age, sex, serum transaminase levels and serum ANA titers were not associated with positivity for serum anti-PD-1 antibody (Table 3). On the other hand, serum levels of anti-PD-1 antibodies were correlated with serum IgG levels ($\rho = 0.37$; $P = 0.002$), but serum IgG levels were not correlated with serum bilirubin levels ($\rho = -0.05$; $P = 0.67$) and prothrombin activities ($\rho = -0.18$; $P = 0.15$). In DILI patients, serum levels of anti-PD-1 antibodies were not correlated with serum IgG levels ($\rho = 0.18$; $P = 0.29$), serum bilirubin levels ($\rho = 0.19$; $P = 0.25$) and prothrombin activities ($\rho = -0.25$; $P = 0.14$).

Histologically, of 71 type 1 AIH patients, seven were diagnosed with acute hepatitis,[11] and the remaining 64 patients were diagnosed with chronic hepatitis. Serum levels of anti-PD-1 antibodies did not differ between patients with acute hepatitis and the others ($P = 0.62$). Positivity for serum anti-PD-1 antibody was shown in 38% (3/8) of patients with acute hepatitis and 38% (24/63) of patients with the chronic disease ($P = 0.97$). Of six cirrhotic patients, three (50%) were positive for serum anti-PD-1 antibody.

Table 2 Clinical features of three DILI patients positive for serum anti-PD-1 antibody

	Case 1	Case 2	Case 3
Age, years	74	38	71
Sex	Female	Female	Female
Causal drug	Allopurinol	Sairei-to	Phenytoin
Laboratory data			
Anti-PD-1 level (OD_{630nm})	0.096	0.180	0.120
WBC (/mm^3)	8300	9500	8800
Hemoglobin (g/dL)	10.3	12.2	10.1
Platelet (×10^4/mm^3)	10.2	30.9	31.8
Bilirubin (mg/dL)	6.2	20.6	7.0
AST (IU/L)	740	1253	204
ALT (IU/L)	340	1263	181
IgG (g/dL)	1.3	1.5	1.5
Prothrombin activity (%)	21	21	46
ANA titer	1:40	1:40	<1:40
Outcome	Died	Survived	Survived

ALT, alanine aminotransferase; ANA, antinuclear antibody; AST, aspartate aminotransferase; DILI, drug-induced liver injury; IgG, immunoglobulin G; OD_{630nm}, optical densities at 630 nm; PD-1, programmed cell death-1; WBC, white blood cell.

Table 3 Associations of positivity for serum anti-PD-1 antibody with clinical features in type 1 autoimmune hepatitis patients

	Positive for anti-PD-1	Negative for anti-PD-1	*P* value
Sample size, *n*	27	44	–
Age, years	59 (29–81)	58 (20–86)	0.70
Sex, female (%)	24 (89%)	39 (89%)	0.97
Revised scoring system proposed by the International Autoimmune Hepatitis Group			
Definite diagnosis	22 (81%)	33 (75%)	0.53
Laboratory data			
WBC (/mm^3)	5600 (1600–7200)	5200 (2300–9300)	0.83
Hemoglobin (g/dL)	13.0 (10.7–15.5)	12.9 (8.7–16.1)	0.43
Platelet (×10^4/mm^3)	15.7 (9.5–24.4)	18.6 (7.4–46.0)	0.054
Bilirubin (mg/dL)	2.7 (0.5–24.4)	1.2 (0.5–24.9)	0.078
AST (IU/L)	346 (30–2466)	255 (39–1561)	0.20
ALT (IU/L)	272 (21–2377)	360 (31–1355)	0.85
IgG (g/dL)	2.8 (1.3–4.9)	2.2 (1.0–4.0)	0.006
Prothrombin activity (%)	69.5 (19.7–102.6)	86.1 (36.0–111.3)	0.002
ANA (%)			0.90
<1:40	1 (4%)	1 (2%)	
≥1:40 and ≤1:80	10 (37%)	18 (41%)	
≥1:160	16 (59%)	25 (57%)	
Liver histology			
Acute hepatitis (%)	3 (11%)	5 (11%)	0.97
Cirrhosis (%)	3 (11%)	3 (7%)	0.53

ALT, alanine aminotransferase; ANA, antinuclear antibody; AST, aspartate aminotransferase; IgG, immunoglobulin G; WBC, white blood cell.

Anti-PD-1 antibody in the diagnosis of type 1 AIH

The ROC analysis showed that measurement of serum levels of anti-PD-1 antibodies was useful for the discrimination of type 1 AIH from DILI (Fig. 2; AUC, 0.80; 95% confidence interval [CI], 0.72–0.89; $P < 0.001$). The AUC of serum IgG levels and ANA titers for the discrimination of type 1 AIH from DILI was 0.86 (95% CI, 0.79–0.94; $P < 0.001$) and 0.83 (95% CI, 0.73–0.93; $P < 0.001$), respectively.

When patients positive for serum anti-PD-1 antibody were diagnosed with type 1 AIH, the sensitivity, specificity, and positive and negative predictive values in the differential diagnosis between type 1 AIH and DILI were 38%, 92%, 90% and 44%, respectively.

In type 1 AIH, five (21%) of 24 patients with serum IgG levels of less than 2 g/dL and nine (41%) of 22 patients with serum ANA titer of 1:40 or less were positive for serum anti-PD-1 antibody. Four (33%) of 12 patients showing serum IgG levels of less than 2 g/dL and serum ANA titer of 1:40 or less were positive for serum anti-PD-1 antibody.

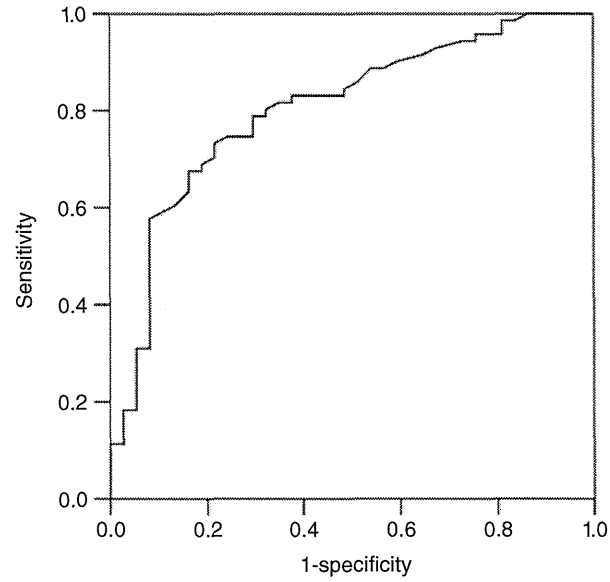

Figure 2 Receiver–operator curve of serum anti-programmed cell death-1 antibodies for the discrimination of type 1 autoimmune hepatitis from drug-induced liver injury. The area under the receiver–operator curve was 0.80 (95% confidence interval, 0.72–0.89; $P < 0.001$).

Anti-PD-1 antibody and initial response to corticosteroid treatment in type 1 AIH

Of 71 type 1 AIH patients, three were not treated with corticosteroid, and the other four were transferred to other hospitals before the normalization of serum ALT levels. So, the association of serum anti-PD-1 antibody with the normalization of serum ALT levels was evaluated in 64 patients treated with corticosteroid. Sixty-two patients (97%) achieved the normalization of serum ALT levels. Cumulative incidences of the normalization of serum ALT levels at 1, 3, 6 and 12 months from the initiation of corticosteroid treatment were 42%, 83%, 94% and 95%, respectively. Patients positive for serum anti-PD-1 antibody achieved later normalization of serum ALT levels than the others (Fig. 3, log-rank test; $P = 0.019$). On the other hand, of the 64 patients treated with corticosteroid, 13 patients initially received i.v. methylprednisolone pulse therapy (125–1000 mg/day for 3 days). Of the remaining 51 type 1 AIH patients initially treated without i.v. methylprednisolone pulse therapy, 16 patients were positive for serum anti-PD-1 antibody and achieved the later normalization of serum ALT levels than the others (log-rank test; $P = 0.048$) although the initial dose of prednisolone (PSL) was similar between those positive for serum anti-PD-1 antibody and the others (40 mg/day [20–60] vs 30 mg/day [15–60]; $P = 0.52$).

In the 64 patients, the univariate Cox proportional hazard model showed that positivity for serum anti-PD-1 antibody, probable diagnosis based on the revised scoring system proposed by the International Autoimmune Hepatitis Group, serum bilirubin levels of 10 mg/dL or more, and prothrombin activity of less than 60% were significantly associated with the later normalization of serum ALT levels. By the multivariate Cox proportional hazard model, positivity for serum anti-PD-1 antibody, probable diagnosis and prothrombin activity of less than 60% were shown to be significantly associated with the later normalization of serum ALT levels (Table 4).

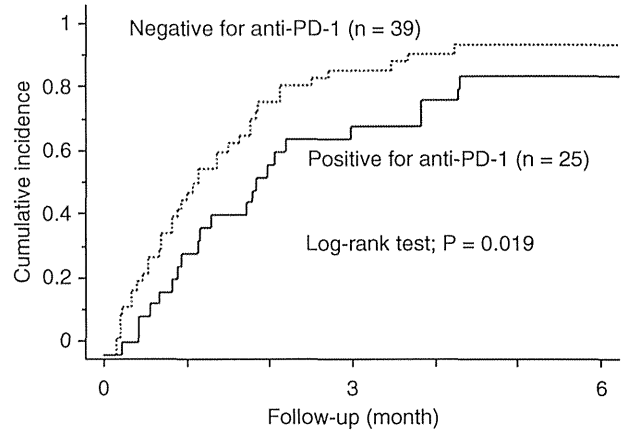

Figure 3 Cumulative incidences of the normalization of serum alanine aminotransferase levels after the initiation of prednisolone treatment. Anti-PD-1, anti-programmed cell death-1.

Table 4 Cox proportional hazard analysis for the factors associated with the later normalization of serum ALT levels

Variables	Univariate		Multivariate	
	Hazard ratio (95% CI)	*P*-value	Hazard ratio (95% CI)	*P*-value
Positivity for anti-PD-1	1.86 (1.10–3.16)	0.022	1.79 (1.03–3.11)	0.038
Age, years	1.01 (0.99–1.02)	0.48	–	–
Sex, male	1.72 (0.73–4.01)	0.21	–	–
Revised scoring system proposed by the International Autoimmune Hepatitis Group				
Probable diagnosis	2.23 (1.20–4.14)	0.011	3.60 (1.84–7.04)	<0.001
Liver histology, acute hepatitis	1.27 (0.57–2.78)	0.57	–	–
Bilirubin, <10 mg/dL	0.53 (0.28–0.96)	0.045	0.99 (0.44–2.20)	0.97
ALT, <300 IU/L	0.90 (0.53–1.51)	0.68	–	–
IgG, <2 g/dL	1.44 (0.83–2.47)	0.19	–	–
Prothrombin activity <60%	2.22 (1.16–4.28)	0.017	3.13 (1.27–7.68)	0.013
ANA <1:80	0.89 (0.52–1.51)	0.66	–	–
Initial PSL dose <40 mg	0.72 (0.43–1.21)	0.22	–	–
Pulse steroid therapy, yes	1.69 (0.90–3.20)	0.11	–	–

ALT, alanine aminotransferase; ANA, antinuclear antibody; CI, confidence interval; IgG, immunoglobulin G; PD-1, programmed cell death-1; PSL, prednisolone.

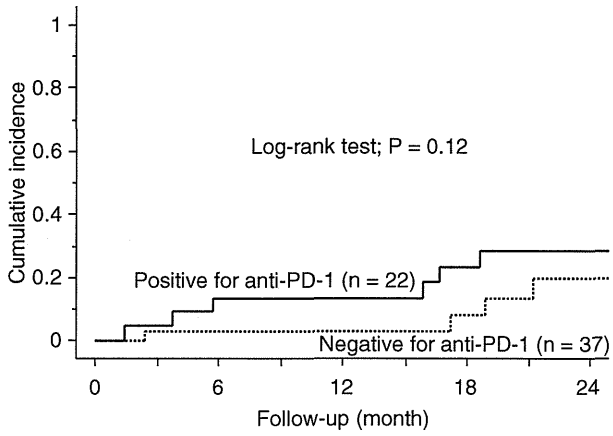

Figure 4 Cumulative incidences of the disease relapse after the normalization of serum alanine aminotransferase levels. Anti-PD-1, anti-programmed cell death-1.

Anti-PD-1 antibody and relapse in type 1 AIH

Of the 62 patients achieving the normalization of serum ALT levels with corticosteroid treatment, 59 were followed up during 1 month or more after the normalization of serum ALT levels. The median follow-up duration was 39 months (2–180). Of these 59 patients, the disease relapse was shown in 12 patients (20%) during the follow up. The disease relapse was shown in eight (36%) of 22 patients positive for serum anti-PD-1 antibody and four (11%) of 37 patients negative for serum anti-PD-1 antibody ($P = 0.018$). The Kaplan–Meier analysis showed that the disease relapse tended to be shown earlier in patients positive for serum anti-PD-1 antibody although the difference did not reach statistical significance (Fig. 4, log-rank test; $P = 0.12$). The duration from the normalization of serum ALT levels to the disease relapse was not different between eight patients positive for serum anti-PD-1 antibody and four patients negative for serum anti-PD-1 antibody (16.3 months [1.5–79.6] vs 18.0 months [2.4–21.2]; $P = 0.81$). Patients showing the disease relapse had higher serum levels of anti-PD-1 antibody (0.096 OD_{630nm} [0.044–0.427] vs 0.064 OD_{630nm} [0.013–0.449]; $P = 0.006$).

DISCUSSION

ELEVATION OF SERUM IgG level and positivity for serum ANA are hallmarks of AIH and main variables included in the scoring systems for the diagnosis of AIH.[1,2] However, almost 15–25% of patients, particularly acute cases, show normal IgG levels at presentation,[12] and 30% were negative for serum ANA.[13] A recent nationwide survey in Japan showed that 40% of type 1 AIH patients showed serum IgG level of 2 g/dL or less, and 10% were negative for serum ANA.[14] In patients showing normal IgG level and/or negativity for serum ANA, the diagnosis of AIH is not always easy. Thus, some markers useful for the diagnosis of AIH are desired.

The purpose of this study was to validate the usefulness of serum anti-PD-1 antibody as an auxiliary diagnostic marker for type 1 AIH. In this study, the prevalence of positivity for serum anti-PD-1 antibody in DILI patients was as low as that of the previous report.[3] On the other hand, the prevalence of positivity for serum anti-PD-1 antibody was not so high in type 1 AIH patients, but the ROC analysis showed that the AUC of serum anti-PD-1 antibody was almost equal to those of serum IgG and ANA. In addition, 21% of type 1 AIH patients showing serum IgG levels of less than 2 g/dL, 41% of those showing serum ANA titer of 1:40 or less, and 33% of those showing serum IgG levels of less than 2 g/dL and serum ANA titer of 1:40 or less were positive for serum anti-PD-1 antibody. Thus, serum anti-PD-1 antibody may be useful for the diagnosis of type 1 AIH as an auxiliary diagnostic marker.

Serum IgG levels have been shown to be associated with the prognosis of type 1 AIH.[15] So, serum IgG of type 1 AIH patients may contain some autoantibodies associated with the disease severity. This study indicated that serum levels of anti-PD-1 antibodies were correlated with the disease severity. This result is consistent with that of the previous report.[3] PD-1 blockade contributes to hyper-responsiveness of $CD8^+$ T cells to antigen and reduced ability of regulatory T cells.[16–18] Serum anti-PD-1 antibodies of type 1 AIH patients may aggravate inflammatory activity through reduced interaction between PD-1 expressed on T cells and its ligands, although further studies are needed. On the other hand, serum IgG levels were not correlated with the disease severity. Thus, not serum IgG but serum anti-PD-1 antibody may be useful as a marker reflecting the disease severity when deciding the treatment strategy.

The previous report has indicated that type 1 AIH patients positive for serum anti-PD-1 antibody achieve the later normalization of serum transaminase levels after the initiation of PSL treatment by the univariate analysis.[3] On the other hand, in this study, positivity for serum anti-PD-1 antibody was confirmed to be associated with the later normalization of serum transaminase levels by the multivariate analysis. Initial response to

corticosteroid treatment has been shown to be a predictive factor for liver-related death or liver transplantation.[19,20] In type 1 AIH patients positive for serum anti-PD-1 antibody, initial treatment should be introduced more carefully.

Repeated relapse is a risk factor for liver-related death or liver transplantation in type 1 AIH.[21,22] In order not to worsen the prognosis, the persistent normalization of serum transaminase levels is important.[23,24] This study indicated that positivity for serum anti-PD-1 antibody was associated with the disease relapse. So, in patients positive for serum anti-PD-1 antibody, the dose reduction of immunosuppressant and the termination of immunosuppressive treatment should be decided more carefully. On the other hand, because of the small sample size, the difference in a cumulative incidence of the disease relapse between patients positive for serum anti-PD-1 antibody and those negative for serum anti-PD-1 antibody did not reach statistical significance when using the Kaplan–Meier method. A further study with a larger sample size and/or longer follow up is needed.

In the previous report,[3] serum levels of anti-PD-1 antibodies were correlated with serum transaminase levels and serum ANA titers. In addition, patients with acute hepatitis showed higher serum levels of anti-PD-1 antibodies than those with the chronic disease. But, these findings were not confirmed by this study.

In conclusion, this study is the first to show that serum anti-PD-1 antibody is a valid auxiliary diagnostic marker for type 1 AIH. Especially, for the diagnosis of type 1 AIH showing normal IgG level and/or negativity for serum ANA, measurement of serum levels of anti-PD-1 antibodies will be useful. In addition, serum anti-PD-1 antibody is confirmed to be associated with the disease severity, response to corticosteroid treatment and the disease relapse. In patients positive for serum anti-PD-1 antibody, the dose reduction of immunosuppressant and the termination of immunosuppressive treatment should be decided more carefully. Hereafter, further studies are needed in order to investigate the functions of anti-PD-1 antibodies in sera of type 1 AIH patients.

ACKNOWLEDGMENTS

THE MINISTRY OF Health, Welfare and Labor of Japan has supported this study as a project of the Intractable Hepato-Biliary Disease Study Group of Japan from April 2011 to March 2014.

REFERENCES

1 Alvarez F, Berg PA, Bianchi FB *et al.* International Autoimmune Hepatitis Group Report: review of criteria for diagnosis of autoimmune hepatitis. *J Hepatol* 1999; **31**: 929–38.
2 Hennes EM, Zeniya M, Czaja AJ *et al.* International Autoimmune Hepatitis Group. Simplified criteria for the diagnosis of autoimmune hepatitis. *Hepatology* 2008; **48**: 169–76.
3 Matsumoto K, Miyake Y, Matsushita H *et al.* Anti-programmed cell death-1 antibody as a new serological marker for type 1 autoimmune hepatitis. *J Gastroenterol Hepatol* 2014; **29**: 110–15.
4 Serriari NE, Gondois-Rey F, Guillaume Y *et al.* B and T lymphocyte attenuator is highly expressed on CMV-specific T cells during infection and regulates their function. *J Immunol* 2010; **185**: 3140–8.
5 Kido M, Watanabe N, Okazaki T *et al.* Fatal autoimmune hepatitis induced by concurrent loss of naturally arising regulatory T cells and PD-1-mediated signaling. *Gastroenterology* 2008; **135**: 1333–43.
6 Maruoka R, Aoki N, Kido M *et al.* Splenectomy prolongs the effects of corticosteroids in mouse models of autoimmune hepatitis. *Gastroenterology* 2013; **145**: 209–20.
7 Topalian SL, Hodi FS, Brahmer JR *et al.* Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med* 2012; **366**: 2443–54.
8 Takikawa H, Onji M. A proposal of the diagnostic scale of drug-induced liver injury. *Hepatol Res* 2005; **32**: 250–1.
9 Takikawa H, Murata Y, Horiike N, Fukui H, Onji M. Drug-induced liver injury in Japan: an analysis of 1676 cases between 1997 and 2006. *Hepatol Res* 2009; **39**: 427–31.
10 Mochida S, Takikawa Y, Nakayama N *et al.* Diagnostic criteria of acute liver failure: a report by the Intractable Hepato-Biliary Diseases Study Group of Japan. *Hepatol Res* 2011; **41**: 805–12.
11 Onji M. Autoimmune Hepatitis Study Group. Proposal of autoimmune hepatitis presenting with acute hepatitis, severe hepatitis and acute liver failure. *Hepatol Res* 2011; **41**: 497.
12 Zachou K, Muratori P, Koukoulis GK *et al.* Review article: autoimmune hepatitis – current management and challenges. *Aliment Pharmacol Ther* 2013; **38**: 887–913.
13 Björnsson E, Talwalkar J, Treeprasertsuk S, Neuhauser M, Lindor K. Patients with typical laboratory features of autoimmune hepatitis rarely need a liver biopsy for diagnosis. *Clin Gastroenterol Hepatol* 2011; **9**: 57–63.
14 Abe M, Mashiba T, Zeniya M, Yamamoto K, Onji M, Tsubouchi H, Autoimmune Hepatitis Study Group-Subgroup of the Intractable Hepato-Biliary Disease Study Group in Japan. Present status of autoimmune hepatitis in Japan: a nationwide survey. *J Gastroenterol* 2011; **46**: 1136–41.

15 Abe M, Onji M, Kawai-Ninomiya K *et al.* Clinicopathologic features of the severe form of acute type 1 autoimmune hepatitis. *Clin Gastroenterol Hepatol* 2007; **5**: 255–8.
16 Velu V, Titanji K, Zhu B *et al.* Enhancing SIV-specific immunity in vivo by PD-1 blockade. *Nature* 2009; **458**: 206–10.
17 McGee HS, Yagita H, Shao Z, Agrawal DK. Programmed Death-1 antibody blocks therapeutic effects of T-regulatory cells in cockroach antigen-induced allergic asthma. *Am J Respir Cell Mol Biol* 2010; **43**: 432–42.
18 Rosenblatt J, Glotzbecker B, Mills H *et al.* PD-1 blockade by CT-011, anti-PD-1 antibody, enhances ex vivo T-cell responses to autologous dendritic cell/myeloma fusion vaccine. *J Immunother* 2011; **34**: 409–18.
19 Czaja AJ. Rapidity of treatment response and outcome in type 1 autoimmune hepatitis. *J Hepatol* 2009; **51**: 161–7.
20 Tan P, Marotta P, Ghent C, Adams P. Early treatment response predicts the need for liver transplantation in autoimmune hepatitis. *Liver Int* 2005; **25**: 728–33.
21 Hoeroldt B, McFarlane E, Dube A *et al.* Long-term outcomes of patients with autoimmune hepatitis managed at a nontransplant center. *Gastroenterology* 2011; **140**: 1980–9.
22 Yoshizawa K, Matsumoto A, Ichijo T *et al.* Long-term outcome of Japanese patients with type 1 autoimmune hepatitis. *Hepatology* 2012; **56**: 668–76.
23 Miyake Y, Iwasaki Y, Terada R *et al.* Persistent normalization of serum alanine aminotransferase levels improves the prognosis of type 1 autoimmune hepatitis. *J Hepatol* 2005; **43**: 951–7.
24 Miyake Y, Iwasaki Y, Terada R *et al.* Persistent elevation of serum alanine aminotransferase levels leads to poor survival and hepatocellular carcinoma development in type 1 autoimmune hepatitis. *Aliment Pharmacol Ther* 2006; **24**: 1197–205.

© 2014. Published by The Company of Biologists Ltd | Development (2014) 141, 2568-2580 doi:10.1242/dev.108548

RESEARCH ARTICLE

STEM CELLS AND REGENERATION

Large hypomethylated domains serve as strong repressive machinery for key developmental genes in vertebrates

Ryohei Nakamura[1], Tatsuya Tsukahara[1], Wei Qu[2], Kazuki Ichikawa[2], Takayoshi Otsuka[1], Katsumi Ogoshi[3], Taro L. Saito[2], Kouji Matsushima[3], Sumio Sugano[4], Shinichi Hashimoto[3,5], Yutaka Suzuki[2], Shinichi Morishita[2,*] and Hiroyuki Takeda[1,*]

ABSTRACT

DNA methylation is a fundamental epigenetic modification in vertebrate genomes and a small fraction of genomic regions is hypomethylated. Previous studies have implicated hypomethylated regions in gene regulation, but their functions in vertebrate development remain elusive. To address this issue, we generated epigenomic profiles that include base-resolution DNA methylomes and histone modification maps from both pluripotent cells and mature organs of medaka fish and compared the profiles with those of human ES cells. We found that a subset of hypomethylated domains harbor H3K27me3 (K27HMDs) and their size positively correlates with the accumulation of H3K27me3. Large K27HMDs are conserved between medaka and human pluripotent cells and predominantly contain promoters of developmental transcription factor genes. These key genes were found to be under strong transcriptional repression, when compared with other developmental genes with smaller K27HMDs. Furthermore, human-specific K27HMDs show an enrichment of neuronal activity-related genes, which suggests a distinct regulation of these genes in medaka and human. In mature organs, some of the large HMDs become shortened by elevated DNA methylation and associate with sustained gene expression. This study highlights the significance of domain size in epigenetic gene regulation. We propose that large K27HMDs play a crucial role in pluripotent cells by strictly repressing key developmental genes, whereas their shortening consolidates long-term gene expression in adult differentiated cells.

KEY WORDS: Epigenetics, DNA methylation, H3K27me3, Vertebrate development, Large hypomethylated domain, Medaka

INTRODUCTION

Cytosine methylation of CpG dinucleotides of the genomic DNA is one of the essential epigenetic modifications in vertebrates. The genomes of all vertebrates studied thus far are globally methylated, and only a small fraction of genomic regions is hypomethylated (Hendrich and Tweedie, 2003; Tweedie et al., 1997). Recent genome-wide analyses have revealed that the majority of gene promoters are hypomethylated (Lister et al., 2009). The hypomethylated promoters are considered as active or permissive for transcription, and DNA methylation at the proximal promoter region is known to be tightly associated with gene silencing (Bird, 2002).

In addition to DNA methylation, histone modifications influence promoter activity (Zhou et al., 2011). Histone H3 lysine 4 (H3K4) methylation is distributed exclusively on hypomethylated DNA regions (Cedar and Bergman, 2009; Hu et al., 2009; Ooi et al., 2007), and positively regulates promoter activity. However, hypomethylated promoters are not exclusively found in genes with active transcription. In embryonic stem cells (ESCs), promoters of developmentally regulated genes are also hypomethylated (Suzuki and Bird, 2008; Xie et al., 2013). These promoters are frequently marked by repressive histone H3 lysine 27 (H3K27) methylation and are proposed to be 'poised' for immediate induction during cell differentiation (Bernstein et al., 2006; Zhao et al., 2007). This poised, but not simply silenced, state of developmental genes is thought to be essential for pluripotent cells to maintain the undifferentiated state with pluripotency. H3K27me3 and hypomethylation at developmental gene promoters were also reported in zebrafish and *Xenopus* early embryos (Akkers et al., 2009; Bogdanovic et al., 2011; Lindeman et al., 2011; Potok et al., 2013; Vastenhouw et al., 2010), suggesting that repression of hypomethylated developmental gene promoters by H3K27me3 is an essential feature in vertebrate development. However, the nature of these hypomethylated promoters marked by H3K27me3 is beginning to be elucidated in the context of development: e.g. the mechanism that regulates the accumulation of H3K27me3.

Recent genome-wide analyses using mammalian cells suggested an antagonistic relationship between the patterns of DNA methylation and H3K27me3 (Brinkman et al., 2012; Lindroth et al., 2008). Importantly, in post-natal mouse brains, DNA methylation at regions flanking proximal promoters was shown to facilitate transcription of neuronal genes by antagonizing H3K27 methylation (Wu et al., 2010). These studies suggest that DNA methylation has diverse functions depending on where it occurs, which in turn indicates the potential role of DNA methylation or hypomethylation outside the proximal promoter in developmental gene regulation. More recently, developmental transcription factor genes were found to be frequently located in conserved large hypomethylated genomic domains, the size of which tends to be much larger than other promoters (Jeong et al., 2014; Long et al., 2013; Xie et al., 2013). However, previous studies have not addressed whether the size of hypomethylated domains affects gene regulation. Furthermore, although global DNA methylation and promoter hypomethylation are conserved among vertebrate genomes, they are not a general feature in other animal species (Suzuki and Bird, 2008). As the role and pattern of DNA methylation vary widely among organisms (Bird, 2002), it is important to ask to what extent their patterns are conserved among vertebrates in order to understand the function and evolution of hypomethylated domains in the vertebrate lineage.

[1]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan. [2]Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa 277-0882, Japan. [3]Department of Molecular Preventive Medicine, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan. [4]Department of Medical Genome, Graduate School of Frontier Sciences, The University of Tokyo, Chiba 277-8561, Japan. [5]Department of Laboratory Medicine, Kanazawa University, Kanazawa 920-8641, Japan.

*Authors for correspondence (moris@cb.k.u-tokyo.ac.jp; htakeda@biol.s.u-tokyo.ac.jp)

Received 29 January 2014; Accepted 9 May 2014

Development (2014) 141, 2568-2580 doi:10.1242/dev.108548

Here, we investigate the epigenomic profiles of medaka blastula embryos (pluripotent cells) and adult tissues by using single-base resolution DNA methylomes from our previous study (Qu et al., 2012) and also by generating genome-wide histone modification maps and additional methylomes. In blastula embryos, large compartmentalized genomic domains with H3K27me3 and DNA hypomethylation (K27HMDs) were found at developmental transcription factor genes. Strikingly, we found that the size of the K27HMD positively correlates with H3K27me3 levels. Thus, whereas hypomethylated promoters usually act as active or permissive domains, K27HMDs provide strong repression when they are broadly hypomethylated. Comparative analyses between medaka and human pluripotent cells revealed both conservation and species-specific changes in K27HMDs, providing insights into the evolution of vertebrate epigenomes. We also found a phenomenon, which we termed 'HMD shortening', at genes with sustained expression in adult tissues. We propose that large K27HMDs play a crucial role in pluripotent cells by strictly repressing key developmental genes, while their shortening consolidates long-term gene expression in adult differentiated cells.

RESULTS

Identification of hypomethylated domains in the medaka genome

Cells from medaka blastula embryos are known to retain pluripotency (Yi et al., 2009). To obtain the global genomic distribution of epigenetic modifications in medaka blastula embryos, we performed ChIP-seq analyses using antibodies against H3K27me3, H3K4me1, H3K4me2, H3K4me3 and H3K27ac (supplementary material Table S1), and integrated these results with previously established medaka single-base DNA methylomes (Qu et al., 2012). At the blastula stage, methylation frequency of individual CpG sites has a clear bimodal distribution (supplementary material Fig. S1A) and thus, contiguous regions of a low level of methylation appeared obvious on a genome browser with sharp boundaries (Fig. 1A). By scanning the whole genome, we identified 15,145 regions with more than nine contiguous low methylated CpG sites (methylation frequency <0.4), termed hypomethylated domains or HMDs (supplementary material Table S4). Gene promoters and CpG islands accounted for the majority of HMDs (supplementary material Fig. S1B). H3K4 methylations exclusively distributed to HMDs, as previously reported (Hu et al., 2009; Ooi et al., 2007), and some parts of those regions are co-enriched with other histone modifications (Fig. 1A).

We noticed that HMDs generally had sharp boundaries (Fig. 1B), suggesting that the boundaries may harbor a specific sequence feature. Because HMDs largely overlap with CpG islands, we analyzed the CpG density around the HMD boundary. We found that the CpG density was significantly higher inside the HMD than outside. Interestingly, the CpG density dropped off around the HMD boundary and a CpG-poor region spanned just outside the HMD (Fig. 1C). CpG-poor regions at HMD boundaries have also been reported in human (Molaro et al., 2011), and may have a specific function in the establishment of the boundaries or may contribute to the sharpness. Furthermore, we identified motifs that are significantly enriched around the boundary sequence (supplementary material Fig. S2). Interestingly, we noticed that one motif showed a high similarity to the CTCF-binding sequence (Fig. 1D). Mapping of CTCF-binding motifs on the medaka genome confirmed that they are indeed highly enriched around HMD boundaries (Fig. 1E). CTCF is known to mediate chromatin looping and functions as a barrier to separate distinct chromatin regions with different modifications (Handoko et al., 2011). Together, these results suggest that HMD boundaries are strictly confined and have specific features in the genomic sequence.

Previous studies reported that nucleosome positioning is affected by CTCF binding and by DNA methylation and histone modifications (Fu et al., 2008; Segal and Widom, 2009; Valouev et al., 2011). To determine the nucleosomal distribution around HMD boundaries bearing a CTCF motif, we re-analyzed the previous data of nucleosome core distribution in medaka blastula embryos (Sasaki et al., 2009). We found clear peaks of nucleosome core signals, and a transition of the periodic pattern of the nucleosomal core position around the CTCF site; the average score of nucleosome core position exhibits a clear 170 bp periodic pattern outside HMDs but the peak becomes low and less defined inside HMDs (Fig. 1F). The binding motif site showed no significant peak, indicating actual CTCF binding at the nucleosome-free region. These results suggest that the nucleosome structure also clearly changes at the HMD boundary from 'packed' to 'loose'.

Large HMDs mark key transcription factor genes for vertebrate development

It is well known that DNA methylation at proximal promoters silences cell type-specific genes, and that permissive promoters are generally hypomethylated (Smith and Meissner, 2013). In some cases, hypomethylated domains expand beyond promoters and cover much larger regions (Bogdanovic et al., 2011; Jeong et al., 2014; Laurent et al., 2010; Long et al., 2013; Xie et al., 2013). However, the significance of such large domain size in gene expression has not been addressed. Interestingly, in medaka genome, although most HMDs were several kb in size and highly enriched with H3K4me2 and other active histone modifications (Fig. 2A,C), we found that a subset of HMDs was extremely large and those regions were frequently enriched with H3K27me3 (Fig. 2B,C). To examine the relationship between the HMD size and H3K27me3, we first classified the HMDs by the existence of H3K27me3 enrichment. 2398 HMDs were found to contain H3K27me3 peaks detected by QuEST software (Valouev et al., 2008) and were classified as K27HMDs (supplementary material Table S4). We also classified the H3K27me3-marked hypomethylated regions using ChromHMM software, which can annotate the regions in a statistically principled manner (Ernst and Kellis, 2012). We confirmed that the K27HMDs largely overlap with the regions identified by ChromHMM (supplementary material Fig. S3). However, regions identified by ChromHMM were frequently divided into smaller fragments, and therefore, we chose K27HMDs that were suitable for the further analysis that focused on domain size. Although H3K27me3-free HMDs (nonK27HMDs) had a high enrichment of H3K4me2, the majority of K27HMDs also had a low but significant enrichment of H3K4me2 (supplementary material Fig. S4).

We next investigated the characteristics of large HMDs. Notably, 12% of K27HMDs were larger than 4 kb, whereas 99.8% of 12,747 nonK27HMDs were less than 4 kb (Fig. 2D). As the majority of HMDs overlapped with promoter regions, we linked medaka genes to promoter-associated HMDs (supplementary material Table S5). Intriguingly, large K27HMD (>4 kb) associated genes (317) and small K27HMDs (<4 kb) associated genes (1295) showed specific features in their functions. Gene ontology analysis showed that terms related to transcription regulation and developmental processes are highly enriched in large K27HMDs (Fig. 2E; supplementary material Fig. S5A). Indeed, 65% of large K27HMD genes encoded DNA-binding factors (supplementary material Table S6), and had crucial functions for embryonic development. By contrast, small K27HMDs showed enrichment of terms for developmental processes, signal transduction and cell communication, with relatively low enrichment

Development (2014) 141, 2568-2580 doi:10.1242/dev.108548

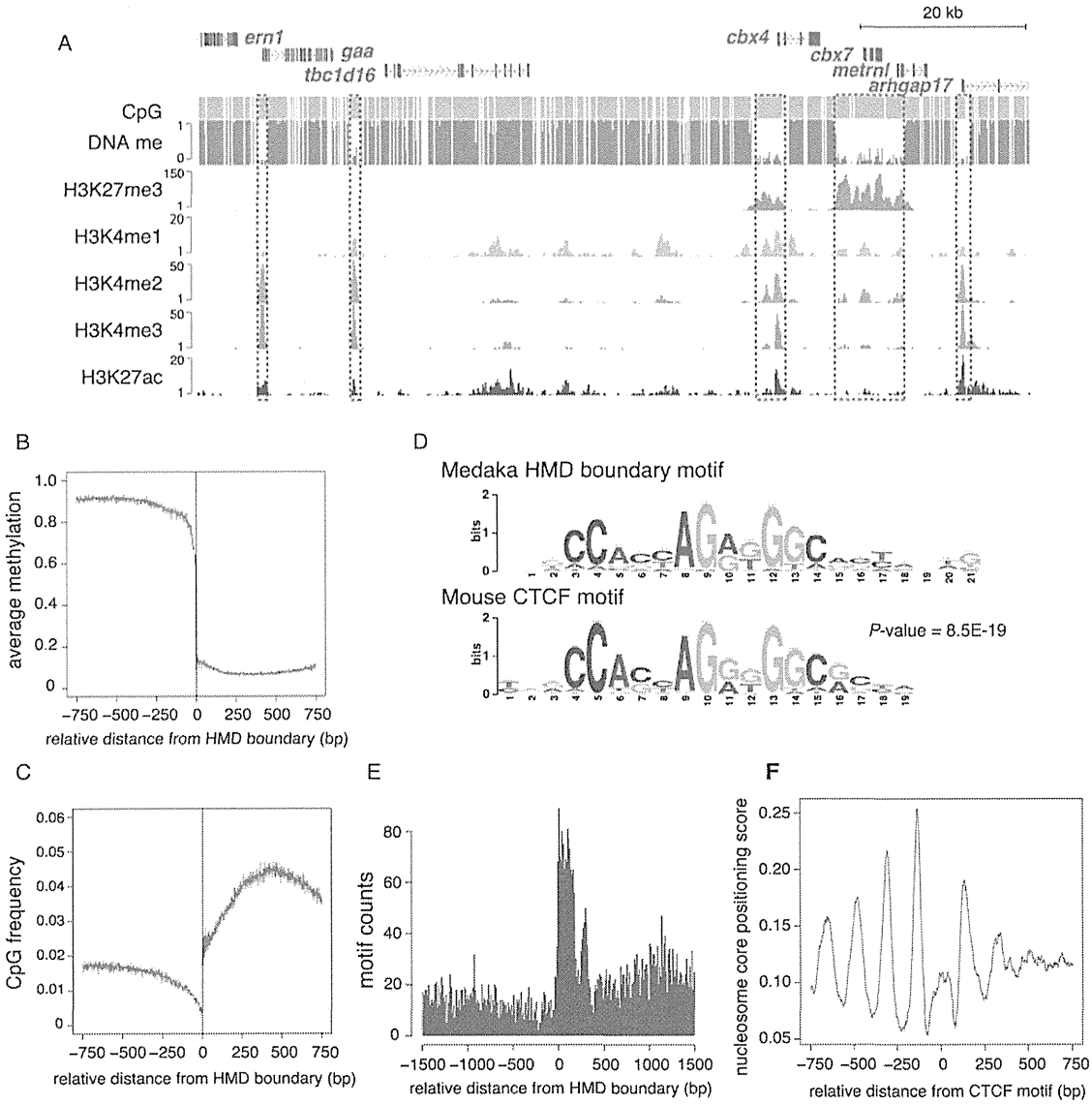

Fig. 1. Identification and characteristics of HMDs in medaka blastula embryos. (A) Genome browser representations of DNA methylation, H3K27me3, H3K4me1, H3K4me2, H3K4me3 and H3K27ac in medaka blastula embryos are shown. Red dashed boxes indicate HMDs. (B) Average methylation ratio around all HMD boundaries. Positions −1 and 1 were excluded as there is no CpG from the definition of the boundary (position 0 is always cytosine of CpG). (C) Frequency of CpG around all HMD boundaries. Positions −1, 0 and 1 were excluded. (D) The DNA motif enriched at HMD boundaries (top) and the mouse CTCF motif (bottom). The *P*-value was determined by the TOMTOM program (Gupta et al., 2007). (E) Distribution of the CTCF motif around boundaries of HMDs larger than 1 kb. Number of CTCF motif centers in 10 bp window were counted. (F) The average local dyad positioning score around boundary-associated CTCF motifs. Downstream regions are hypomethylated.

for transcription factors (Fig. 2E; supplementary material Fig. S5A). Thus, the large K27HMDs mark a specific set of developmental genes, and may have an important role in vertebrate development. We also confirmed that the HMD size does not reflect gene length, as large K27HMDs tended to have shorter genes than small ones (supplementary material Fig. S5B).

It is known that the teleost underwent whole-genome duplication (Jaillon et al., 2004; Kasahara et al., 2007; Taylor et al., 2003), and we found that some duplicated genes encoding transcription factors showed different HMD size. We picked two duplicated gene pairs [*pax6* (ENSORLG00000009913 and ENSORLG00000000847) and *tbx2* (ENSORLG00000014792 and ENSORLG00000010011)] for expression analysis because, for each pair, one gene is associated with a large K27HMD and the other with a small K27HMD (supplementary material Fig. S5C,D). *In situ* hybridization of those gene pairs in medaka embryos demonstrated that the genes with large HMDs are expressed in a tissue-specific pattern (supplementary material Fig. S5E,F), which is nearly identical to the conserved expression pattern in vertebrates, including mouse (Harrelson et al., 2004; Puschel et al., 1992). By contrast, the genes with small HMDs showed no or partial expression of the conserved pattern (supplementary material Fig. S5E,F). These results suggest a conserved function for large K27HMD genes, and strengthen the possibility that the size of the HMD reflects the gene function.

The size of K27HMD and the number of CpGs in the domain correlate with H3K27me3 levels

We assumed that the size of the HMDs contributes to the transcriptional regulation of developmental transcription factor

Development (2014) 141, 2568-2580 doi:10.1242/dev.108548

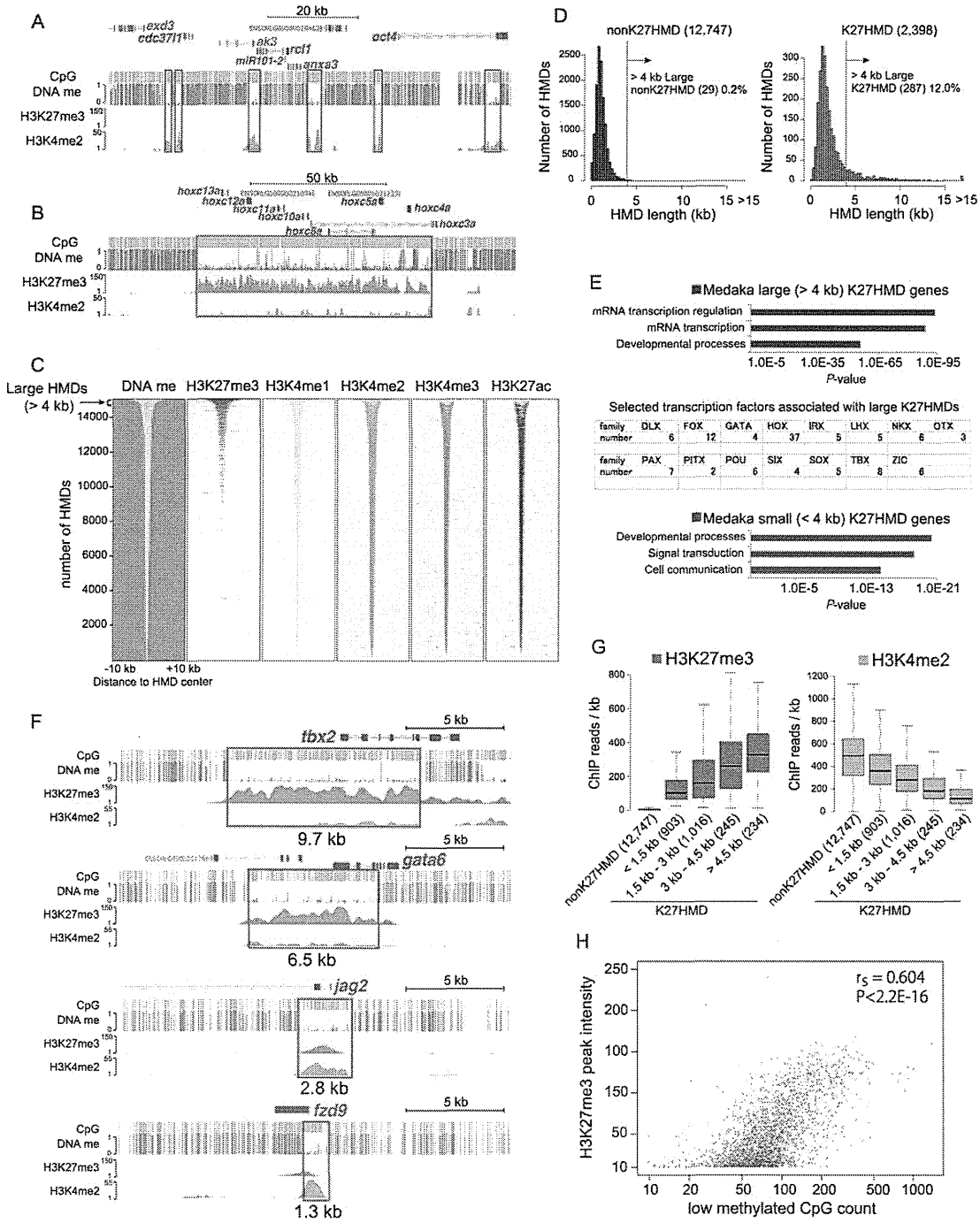

Fig. 2. Large HMDs have high H3K27me3 enrichment and mark key transcription factor genes. (A,B) Genome browser representations of DNA methylation, H3K27me3 and H3K4me2 in medaka blastula embryos are shown for representative nonK27HMD (A) and large K27HMD (B). (C) Heat-map representations of DNA methylation, H3K27me3, H3K4me1, H3K4me2, H3K4me3 and H3K27ac are shown for all HMDs in medaka blastula embryos. HMDs are ordered by size. Dark colors represent high signal; white represents low signal. (D) Size distributions of nonK27HMDs and K27HMDs in blastula embryos. (E) Functional annotations of genes marked by large (top) and small (bottom) K27HMDs. The top three over-represented Gene Ontology (GO) PANTHER biological process terms are shown. The values of *x* axes (in logarithmic scale) correspond to the *P* values calculated by the DAVID tool (Huang et al., 2009). (F) Genome browser representations of DNA methylation, H3K27me3 and H3K4me2 in medaka blastula embryo are shown for large K27HMDs (*tbx2* and *gata6*; top two) and small K27HMDs (*jag2* and *fzd9*; bottom two). (G) Boxplots show the correlation between HMD length and ChIP reads per kb for indicated modifications. In the box plots, bottom and top of the boxes correspond to the 25th and 75th percentiles; the internal band is the 50th percentile (median). The plot whiskers extending outside the boxes correspond to the lowest and highest datum within 1.5 interquartile ranges of the lower and upper quartiles, respectively. (H) Comparison of the number of low methylated CpG sites and the highest H3K27me3 ChIP peak intensity inside K27HMD. Spearman's rank correlation coefficient (r_s) and *P* values are shown.

genes. The fact that most of the large HMDs are enriched with H3K27me3 led us to hypothesize that the size of the HMD is one of the determinants of the level of H3K27me3 accumulation. Comparison of K27HMD size and ChIP signal enrichment or ChIP peak intensity revealed a significant correlation (Fig. 2F,G; supplementary material Fig. S6A,C). By contrast, negative correlation was found between

Development (2014) 141, 2568-2580 doi:10.1242/dev.108548

domain size and active modification levels (H3K4me1, H3K4me2, H3K4me3 and H3K27ac). Among those active modifications, the negative correlation was most significant for H3K4me2 (Fig. 2F,G; supplementary material Fig. S6A,B). These results indicate that larger K27HMDs have a stronger repressive property for transcription. Given that large K27HMD genes mostly encode transcription factors that are crucial for development, the strong repression of those genes in blastula embryos might prevent improper cell differentiation and maintain stemness.

Previously, Polycomb group (PcG) proteins that mediate H3K27me3 accumulation were reported to preferentially bind to CpG islands (Deaton and Bird, 2011; Ku et al., 2008; Woo et al., 2010), but DNA methylation reduces this binding (Hagarman et al., 2013; Wu et al., 2010). We therefore counted the number of low methylated CpG sites inside the K27HMDs and examined their relationship with H3K27me3 levels. A significant correlation was observed between the low methylated CpG count and the H3K27me3 ChIP peak intensity (Fig. 2H), supporting the possibility that a large K27HMD provides a large number of unmethylated CpGs that can potentially recruit PcG proteins.

Comparison of HMDs between medaka and human pluripotent stem cells

To test whether the importance of the HMD size holds true for other vertebrates, we applied our analysis pipeline to a whole-genome dataset from human ESCs (hESCs) (Lister et al., 2009, 2011). hESCs were chosen because we assumed that they are equivalent to medaka blastula cells in terms of pluripotency. We first tested whether the HMD boundaries in hESCs are also enriched with CTCF. Mapping of CTCF ChIP-seq peaks from the ENCODE Project Consortium (2011) around HMDs showed the highest enrichment at the HMD boundary (supplementary material Fig. S7), suggesting a conserved feature of the HMD boundary between medaka and human. Next, we examined the epigenetic status at promoters of orthologous genes and its conservation. All medaka genes annotated to single human orthologous genes in Ensembl database (13,301 genes) were used for comparison between the two species. We first found that the majority of medaka promoters (59%) have similar DNA and histone modifications in human ES cells (Pearson's Chi-squared test $P<2.2E-16$; Fig. 3A; supplementary material Table S7). Furthermore, the comparison of HMD size between medaka and human further revealed a conserved tendency: whereas nonK27HMDs largely reside in small size fraction in both medaka and human (Fig. 3B), genes marked by large K27HMDs in medaka embryos are also marked by large K27HMDs in hESCs (Fig. 3B,C). We defined HMDs larger than 8 kb as large HMDs in human, so that the number of genes designated as large K27HMDs is equal to large K27HMD (>4 kb) genes in medaka. We confirmed that the genes associated with conserved large K27HMDs are highly enriched for GO terms related to transcription regulation and developmental processes (Fig. 3D). Furthermore, at the level of protein sequence, genes found in large K27HMDs are more conserved between the two organisms than those in small K27HMDs or nonK27HMDs, and genes with no HMD (i.e. methylated) showed the lowest conservation (supplementary material Fig. S8), indicating that large K27HMD genes are under strong evolutional constraint.

As in medaka blastula embryos, large K27HMDs in hESCs accumulated higher levels of H3K27me3 (Fig. 3E), and the correlation between the number of low methylated CpG sites and H3K27me3 levels was also significant in hESCs (supplementary material Fig. S9). Thus, large K27HMDs seem to be more repressive than small K27HMDs. We tested this idea by analyzing the transcriptome data set from a previous study (Lister et al., 2011). First, as expected, nonK27HMD genes exhibit high levels of expression, whereas the majority of no HMD (methylated) genes are silenced (Fig. 3F). Consistent with the poised model (Bernstein et al., 2006; Pan et al., 2007), the expression levels of K27HMD genes tend to be low. Importantly, large K27HMD genes showed significantly lower expression than small ones, almost similar levels to methylated genes (Fig. 3F). This tendency was also observed in blastula embryos, but was not so evident (data not shown), probably owing to the presence of maternally derived transcripts and due to unusual transcriptional environment of mid-blastula transition that the blastula embryos had just experienced (Aizawa et al., 2003). Taken together, these results suggest the large size of K27HMD at key transcription factor genes is conserved among vertebrate species, and have strong repressive effect on gene expression in pluripotent cells.

Human-specific K27HMDs mark genes related to neuronal activities

Although the epigenetic status of homologous genes is largely conserved between medaka and human pluripotent cells, a subset of genes has been subjected to differential DNA and histone methylations. We looked into these differences focusing on human-specific K27HMD genes, because they could reflect changes in function that occurred during vertebrate evolution. For this purpose, K27HMD genes in hESCs were classified into three categories according to the epigenetic status of their medaka counterparts. Class 0 is a category of K27HMD genes shared by the two species as described above, whereas genes of the latter two became K27HMDs in human from nonK27HMDs (Class I) or methylated (Class II) in medaka, respectively (Fig. 3A). We then determined GO terms enriched for these three classes of hESC K27HMDs. We first confirmed that developmental genes are highly enriched in Class 0 K27HMDs (Fig. 4A). By contrast, no such high enrichment for developmental genes was observed for class I and class II genes. Furthermore, the large K27HMDs were mostly included in Class 0 (supplementary material Fig. S10), strongly suggesting that this epigenetic machinery is essential for development and thus conserved among vertebrates. Interestingly, class II genes are more closely associated with signal transduction and neuronal activities (Fig. 4B,C; supplementary material Table S8), whereas Class I genes did not show any enrichment of specific terms. We also failed to find any preference of GO terms for genes in medaka-specific K27HMDs (data not shown).

The human-specific HMDs could be due to a different differentiation state between the two types of cells, medaka blastula cells and hESCs, despite their presumed pluripotency. However, sequence conservation of three classes of human K27HMDs among vertebrates revealed that Class I and II HMDs are more divergent than Class 0 (Fig. 4D), suggesting that genetic variations in cis-elements, rather than cellular characteristics of the cells, account for the observed differences in epigenetic modifications between medaka and human.

Dynamics of the epigenetic state of *zic1*/*zic4* genes during development

Next, we sought to determine how the epigenetic state of large K27HMDs changes from embryonic stages to adult. We focused on the *zic1* and *zic4* (*zic1*/*zic4*) genes because they have a particularly large K27HMD (26 kb) in medaka blastula embryos (Fig. 5A) and their function has been extensively studied in medaka (Kawanishi et al., 2013; Moriyama et al., 2012). *zic1*/

Development (2014) 141, 2568-2580 doi:10.1242/dev.108548

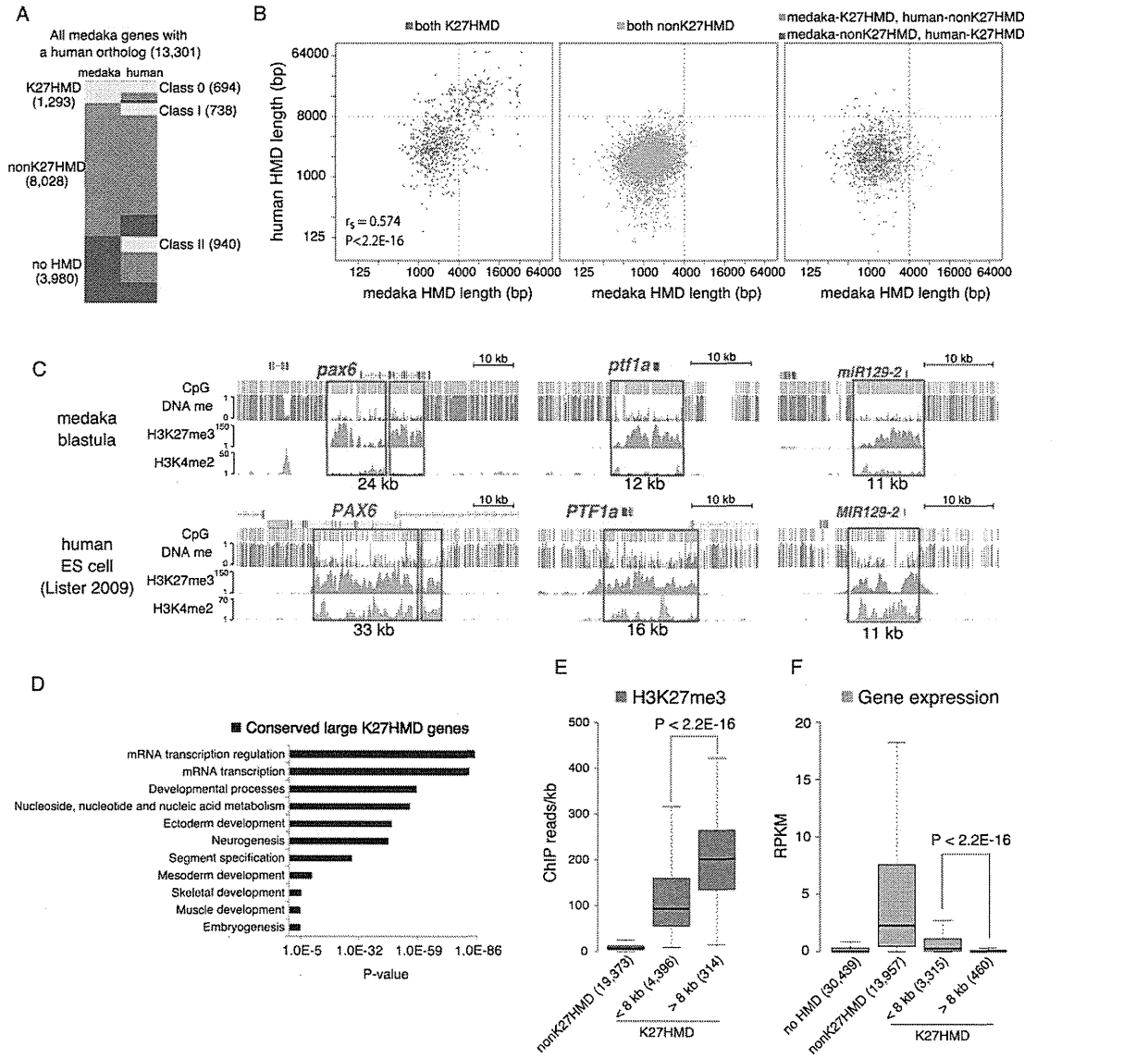

Fig. 3. Comparison of HMDs between medaka and human. (A) HMD status in medaka blastula embryos and human ES cells is shown for promoters of 13,301 genes that had an orthologous human gene annotated in the Ensembl database. Yellow indicates the presence of a K27HMD; orange indicates nonK27HMD; red indicates no HMD at promoter regions. (B) Comparison of the size of HMDs associating with gene promoters for medaka gene and its human ortholog. The HMD size correlation is shown in each panel for genes marked by the same type of HMDs in medaka and human (left, K27HMD; middle, nonK27HMD) and different types of HMDs (right). Spearman's rank correlation coefficient (r_s) and *P* value are shown for genes with conserved K27HMDs (left panel). (C) Genome browser representations of DNA methylation, H3K27me3 and H3K4me2 in medaka (top) and human (bottom) are shown for conserved large K27HMDs (*pax6*, *ptf1a* and *miR129-2*). (D) Functional annotation of genes marked by conserved large (>4 kb for medaka and >8 kb for human) K27HMDs. The over-represented GO PANTHER biological process terms are shown. The values of *x* axes (in logarithmic scale) correspond to the *P* values calculated by the DAVID tool. (E,F) Boxplots show H3K27me3 ChIP reads per kb (E) and gene expression levels (F) in human ES cells for each HMD category. The number of HMDs (E) and genes linked to each HMD category (F) are shown under the boxes. *P* values were calculated using non-paired Wilcoxon tests. In the box plots, the bottom and top of the boxes correspond to the 25th and 75th percentiles; the internal band is the 50th percentile (median). The plot whiskers extending outside the boxes correspond to the lowest and highest datum within 1.5 interquartile ranges of the lower and upper quartiles, respectively.

zic4, organized head-to-head with a small intergenic sequence (~3.4 kb), encode zinc-finger transcription factors and function in the neuronal development and specification of dorsal fate in the trunk (Kawanishi et al., 2013; Merzdorf, 2007). They are thought to share cis-regulatory elements and thus their expression is nearly identical. It is known that their expression is activated by secreted factors during embryogenesis, and is autonomously maintained throughout life in the dorsal parts of somite derivatives, i.e. the myotome, dermis and vertebrae (Kawanishi et al., 2013). Thus, these genes served as a good example with which to analyze the change in epigenetic states of key developmental genes from embryo to adult. We isolated dorsal (*zic*-positive) and ventral (*zic*-negative) myotome separately from adult fish and generated methylomes and histone modification maps for each half (supplementary material Tables S1 and S3, Fig. S11).

Although the overall methylation pattern was similar between dorsal and ventral myotome, we noticed a significant difference in the chromatin state at the *zic1/zic4* locus (Fig. 5A). In the dorsal myotome, the *zic1/zic4* locus showed low H3K27me3 levels and high H3K4me2 levels, consistent with the state of active transcription. Surprisingly, however, we found large blocks of DNA hypermethylation at regions outside the H3K4me2-enriched promoter regions, which led to the

Development (2014) 141, 2568-2580 doi:10.1242/dev.108548

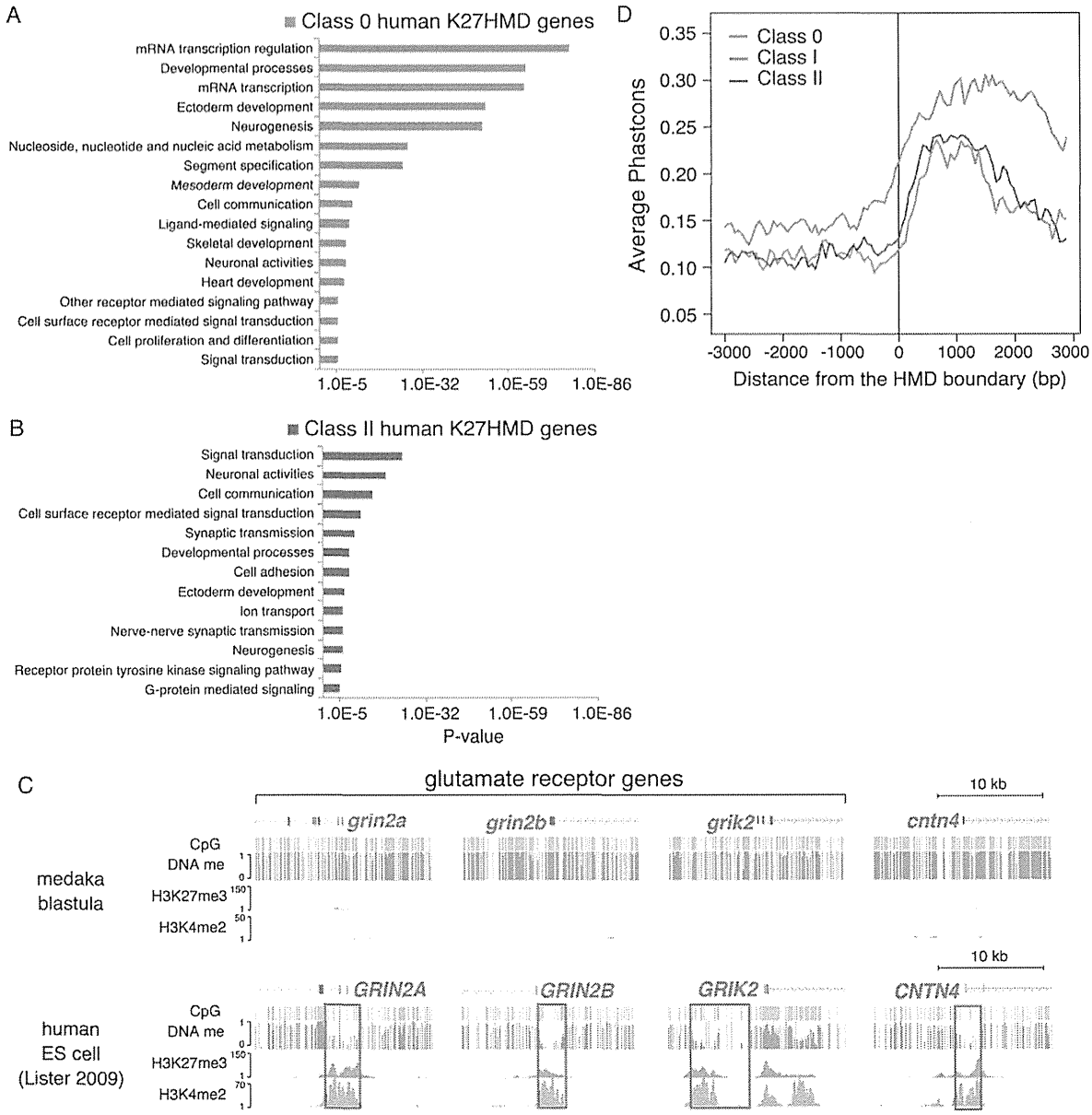

Fig. 4. Characteristics of human-specific K27HMDs. (A,B) Functional annotation of genes marked by Class 0 (K27HMD in medaka; A) and Class II (methylated in medaka; B) human K27HMDs. The top over-represented GO PANTHER biological process terms are shown. The values of *x* axes (in logarithmic scale) correspond to the *P* values calculated by the DAVID tool. (C) Genome browser representations of DNA methylation, H3K27me3 and H3K4me2 in medaka (top) and human (bottom) are shown for Class II genes (*grin2a*, *grin2b*, *grik2* and *cntn4*). (D) Average vertebrate PhastCons profiles around the boundaries of Class 0 (K27HMD in medaka; green), Class I (nonK27HMD in medaka; orange) and Class II (methylated in medaka; red) human K27HMDs.

shortened HMD around the promoter region (Fig. 5A, blue box). We termed this phenomenon 'HMD shortening'. By contrast, the ventral myotome maintained the blastula-type epigenetic state: high H3K27me3 and low DNA methylation. Given that HMD size correlates with transcriptional repression level, this HMD shortening might promote *zic1/zic4* expression. To examine when this HMD shortening occurs during development, we additionally investigated the DNA methylation pattern in dorsal and ventral myotome at the hatching stage, a stage when the dorsal-specific expression of *zic1/zic4* is already induced (Kawanishi et al., 2013). As expected, ChIP-qPCR revealed that the pattern of active and repressive histone modifications was already established in the dorsal and ventral myotome by the hatching stage (Fig. 5B; supplementary material Fig. S12A). By contrast, the dorsal-specific DNA hypermethylation was not detected at this stage (Fig. 5A). Bisulfite sequencing at the larval stages revealed a gradual increase in DNA methylation as larval development progressed (Fig. 5C). Thus, HMD shortening occurs after the establishment of active histone modifications and gene expression in dorsal myotome. These results suggest that DNA hypermethylation in the *zic1/zic4* HMD is not necessary for the initial induction but rather for the maintenance of gene expression.

We hypothesized that HMD shortening at later stages depends on the active chromatin state. To explore this possibility, we determined the DNA methylation pattern in the *zic1/zic4* mutant *Double anal fin* (*Da*). The *Da* mutant has a large transposon insertion that impairs the mesoderm enhancer of *zic1/zic4*, leading to a dramatic reduction in myotome expression and ventralized dorsal structures in the trunk (supplementary material Fig. S12A; Moriyama et al., 2012). ChIP-qPCR analysis confirmed the high enrichment of H3K27me3 remained in the dorsal myotome of *Da*

Development (2014) 141, 2568-2580 doi:10.1242/dev.108548

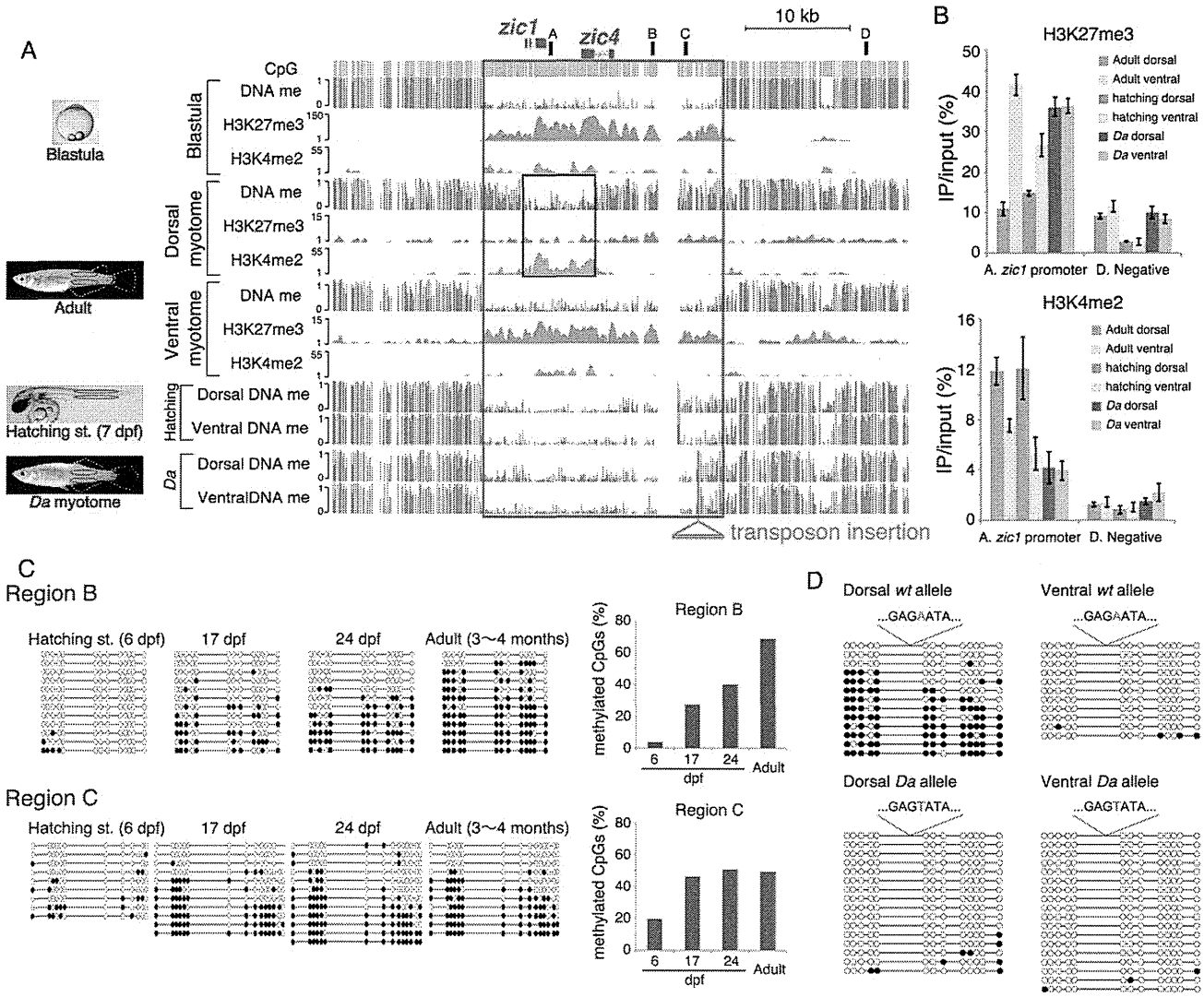

Fig. 5. Active chromatin-dependent K27HMD shortening at *zic1/zic4* locus in adult myotome. (A) Genome browser representation at the *zic1/zic4* loci. DNA methylation, H3K27me3 and H3K4me2 are shown for blastula embryos, and for adult dorsal and ventral myotome; DNA methylation is shown for dorsal and ventral myotome of hatching stage larvae and adult *Da* mutant. Red box indicates a large K27HMD identified in the blastula embryo and blue box indicates a shortened HMD in adult dorsal myotome. Navy bars (labeled A-D) indicate regions examined for ChIP-qPCR and bisulfite sequencing. (B) ChIP-qPCR analysis of dorsal and ventral myotome from *wt* adult, hatching stage larvae and *Da* adult for H3K27me3 (top) and H3K4me2 (bottom) at regions indicated in A. Error bars represent s.d. from three technical replicates. (C) Bisulfite sequencing in dorsal myotome at multiple stages after hatching. Graphs indicate the ratio of methylated CpGs at each stage. The examined regions are indicated in A. (D) Bisulfite sequencing in adult *Da* heterozygous fish at region C in A. Even in the same nucleus, only the *wt* allele shows DNA hypermethylation.

(Fig. 5B; supplementary material Fig. S12B). Strikingly, bisulfite-seq showed a complete absence of the dorsal-specific DNA hypermethylation in myotome of *Da* (Fig. 5A), suggesting the requirement of *zic1/zic4* activation for HMD shortening. To further elaborate this possibility, we made *Da* heterozygous fish and investigated DNA methylation of each allele. *wt* and *Da* alleles were distinguished by single base nucleotide variations. In the dorsal myotome of heterozygous fish, only the *zic1/zic4* locus of the *wt* allele is activated and that of the *Da* allele remained repressive because of the transposon insertion. Bisulfite sequencing revealed that the *wt* allele was highly methylated, whereas the *Da* allele remained hypomethylated (Fig. 5D). Thus, we conclude that the activation of the *zic1/zic4* locus, but not of Zic1/Zic4 proteins or their downstream factors, induces HMD shortening. Collectively, these results suggest that once *zic1/zic4* are induced during embryogenesis, they are autonomously subject to HMD shortening during growth.

HMD shortening associates with sustained gene expression in matured organs

The epigenetic profile of the *zic1/zic4* locus suggested that large K27HMDs have a repressive function, while the shortening of the K27HMD associates with their sustained expression. We also noticed that the large K27HMD at Hox gene clusters underwent the remarkable DNA hypermethylation that led to shortened HMDs at the active promoter in adult myotome (supplementary material Fig. S12C). To address whether HMD shortening at active gene loci is a general feature in adult tissues, we compared the epigenomes of the adult dorsal myotome and liver with that of blastula embryos. For liver epigenome data, we generated histone modification maps (supplementary material Table S1), and integrated these with methylome data from a previous study (Qu et al., 2012). We found that the majority of K27HMDs have unchanged methylation levels in adult tissues, but a significant

Development (2014) 141, 2568-2580 doi:10.1242/dev.108548

proportion of HMDs were subjected to DNA hypermethylation in which more than 5% of their CpG sites became highly methylated (>0.4) (Fig. 6A; supplementary material Tables S4 and S5). Indeed, 8.3% of K27HMDs identified at the blastula stage acquired elevated DNA methylation in both adult myotome and liver, and 9.6% and 5.3% showed elevated DNA methylation specifically in adult myotome and adult liver, respectively (Fig. 6B). Large K27HMDs (>4 kb) tend to be more frequently methylated in adult tissues, 15.7% for both, and 20.9% and 5.2% only for myotome and liver, respectively (Fig. 6B). Strikingly, the large K27HMDs with elevated DNA methylation still retained low methylation levels around TSSs in adult tissues, resulting in the HMD shortening toward the promoter region (Fig. 6A,C; supplementary material Fig. S13A). Similar to *zic1/zic4*, this change in DNA methylation occurred mainly after the hatching stage in myotome (Fig. 6C).

Consistent with the fact that larger K27HMD have higher H3K27me3 enrichment in pluripotent cells, shortened K27HMD in adult tissues showed significantly lower levels of H3K27me3 than unchanged K27HMDs (Fig. 6D; supplementary material Fig. S13B). Lower levels of H3K27me3 associated with elevated

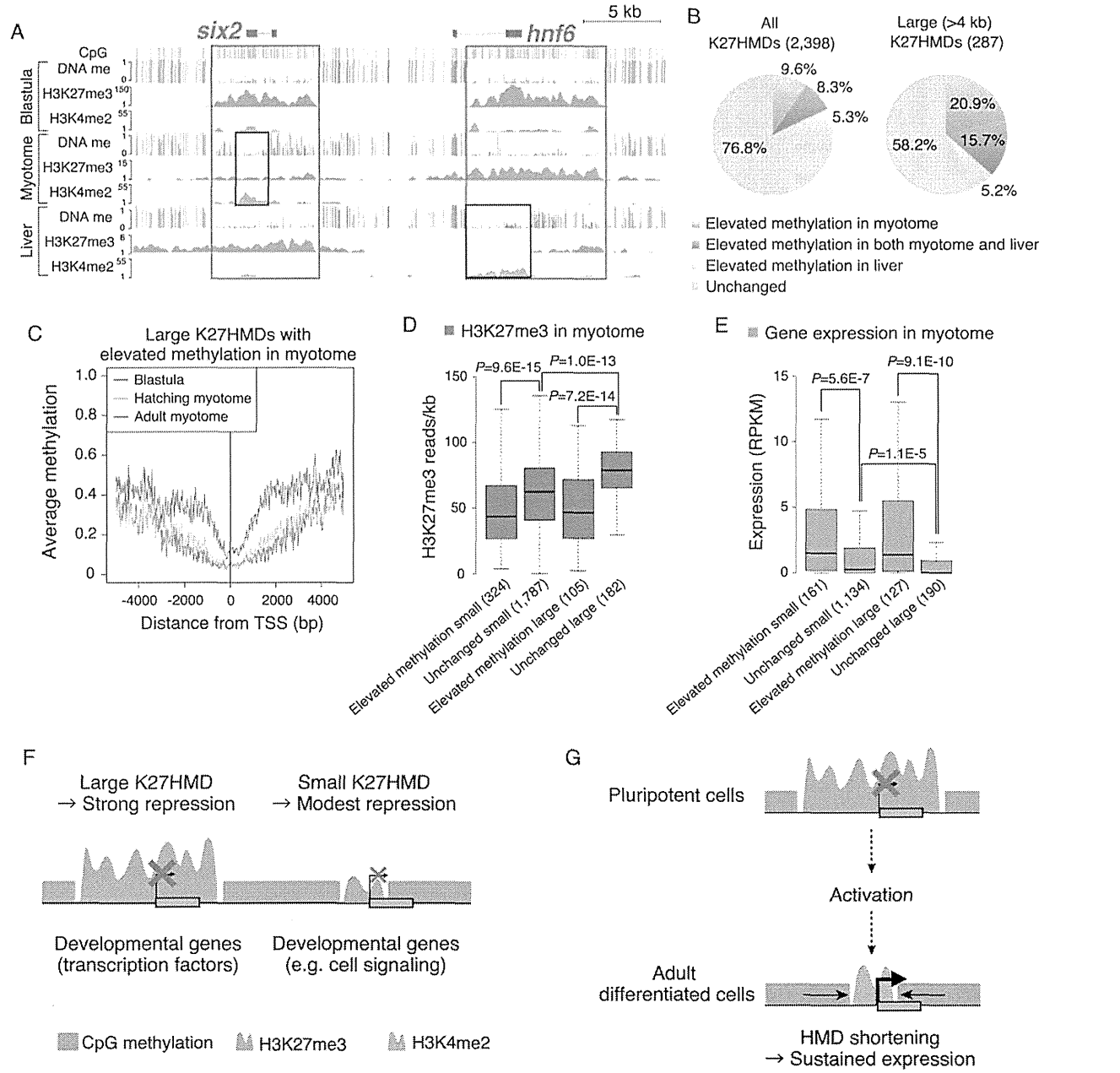

Fig. 6. Large K27HMD regions acquire non-promoter DNA methylation at active loci. (A) Genome browser representation of DNA methylation, H3K27me3 and H3K4me2 enrichment are shown for K27HMD with myotome-specific hypermethylation (*six2*; left) and liver-specific hypermethylation (*hnf6*; right). Red box indicates large K27HMD identified in the blastula embryo and blue box indicates shortened HMD in adult dorsal myotome. (B) Fraction of K27HMDs with DNA hypermethylation in adult myotome and liver for all K27HMD (left) and large K27HMD (right). (C) Average DNA methylation around TSSs marked by a large K27HMD with elevated DNA methylation in myotome at multiple stages. (D,E) Boxplots show H3K27me3 enrichment (D) and gene expression (E) at hypermethylated (elevated methylation) and unchanged K27HMDs in adult myotome. The number of HMDs (E) and genes linked to each HMD category (F) are shown under the boxes. *P* values were calculated using non-paired Wilcoxon tests. (F,G) Models for the HMD size-dependent repression of developmental genes in pluripotent cells (F) and for HMD shortening associated with sustained gene expression in adult differentiated cells (G).

Development (2014) 141, 2568-2580 doi:10.1242/dev.108548

DNA methylation because the levels of H3K27me3 accumulation in blastula embryos did not show any difference between the two groups (elevated methylation and unchanged methylation in adult) of K27HMDs (supplementary material Fig. S13D). To examine whether HMD shortening is correlated with gene expression, we performed RNA-seq for adult myotome and liver (supplementary material Table S2). Consistent with the reduced H3K27me3 levels, gene expression levels were significantly higher in the K27HMDs with elevated methylation than in K27HMDs with unchanged methylation (Fig. 6E; supplementary material Fig. S13C). Thus, K27HMD shortening is associated with reduced H3K27me3 levels and with active gene expression. By contrast, DNA methylation occurred in some nonK27HMDs and, in this case, a negative correlation between DNA methylation and gene expression was observed (supplementary material Fig. S14). These results suggest opposing functions of DNA methylation on the expression of nonK27HMDs and K27HMDs.

Importantly, a correlation between the HMD size and H3K27me3 levels was also observed for unchanged K27HMDs in adult myotome and liver (Fig. 6D; supplementary material Fig. S13B, unchanged small versus unchanged large). Furthermore, the expression level was significantly lower for unchanged large K27HMDs than for small K27HMDs (Fig. 6E; supplementary material Fig. S13C). Taken together, these data suggest that the large K27HMDs have a strong repressive effect on gene expression in pluripotent and adult matured cells (Fig. 6F), and the shortening of the large K27HMDs may facilitate persistent gene expression in adult tissues (Fig. 6G).

To further investigate whether HMD shortening occurs in other vertebrate species, we analyzed methylome and transcriptome data from zebrafish sphere stage embryo (an equivalent stage for medaka blastula) and adult myotome (Potok et al., 2013). We identified 237 HMDs larger than 8 kb in zebrafish sphere embryos, and 139 of those showed elevated methylation in adult myotome (supplementary material Table S9). Expression levels of genes associated with elevated methylation were significantly higher than those of unchanged HMD genes (supplementary material Fig. S15). These results suggest that HMD shortening also associates with active gene expression in zebrafish. Therefore, the regulation of HMD size could be an important epigenetic mechanism for the strict and stable regulation of key developmental genes in vertebrates.

DISCUSSION

Previous studies have demonstrated that the H3K27me3 domains often occupy several kb around promoters of developmental genes, but at some gene loci, the domains exceed this to cover much larger regions (Zhao et al., 2007). Very recent studies also identified the large genomic domains with DNA hypomethylation at transcription factor gene loci in various vertebrate species (Jeong et al., 2014; Long et al., 2013; Xie et al., 2013). However, the biological significance of the size of these genomic domains has not been addressed. H3K27me3 marked promoters are proposed to keep developmental genes poised for immediate activation in pluripotent cells (Bernstein et al., 2006). Indeed, the transcriptional activity of K27HMD genes in undifferentiated cells was found to be low, but slightly higher than that of methylated genes (Fig. 3F, small K27HMD genes versus methylated genes), suggesting leaky transcription under polycomb-mediated repression. Importantly, we found that the H3K27me3 level correlates with K27HMD size and the transcription level of genes in large K27HMDs is significantly lower than that in small ones in both differentiated medaka tissues and human ESCs. In pluripotent cells, the large K27HMDs preferentially mark transcription factor genes with crucial function in development, whereas the smaller ones tend to mark genes related with signal transduction. Transcription factors in large K27HMDs, therefore, are strictly shut off if not required. A previous study in zebrafish sperm also reported that genes with high levels of H3K27me3 at proximal promoters mostly encode transcription factors (Wu et al., 2011), which are largely overlapped with genes identified by the domain size (large K27HMD genes) in our study. This further emphasizes the generality of size-dependent H3K27me3 accumulation and strong repression of transcription factor genes in vertebrates. The strict repression is important presumably because the derepression of those transcription factors would result in inappropriate cell differentiation (Boyer et al., 2006; Fujikura et al., 2002) or malignancies such as cancer (Darnell, 2002). The precise mechanism regulating the amount of H3K27me3 awaits future studies. However, our data suggest that large HMDs provide a broad platform with unmethylated CpGs that potentially recruit PcG proteins to enrich H3K27me3, while small HMDs have limited capability to bind PcG proteins.

Recent studies have identified the super-enhancer in which large domains occupied by clusters of enhancers drive strong expression of key cell-identity genes (Whyte et al., 2013). The concept of the large K27HMD is reminiscent of that of the super-enhancer, but the two large genomic domains have opposing functions: robust repression versus strong activation. These large domains with dense epigenetic modifications may strengthen the repression or activation effect, and cooperatively contribute to the control of cell identity.

During growth and differentiation, large K27HMDs undergo substantial changes in histone modification and DNA hypermethylation when their genes are activated. Interestingly, Polycomb-target genes in ESCs often acquire hypermethylation after differentiation (Mohn et al., 2008; Rush et al., 2009), but the underlying mechanism is unknown (Deaton and Bird, 2011). The analysis of the large K27HMD harboring the *zic1/zic4* genes revealed that DNA hypermethylation is locus activity dependent and gradually occurs after the histone modification pattern has changed. A likely scenario is that the reduction of H3K27me3 allows the accumulation of DNA methylation outside a proximal promoter harboring H3K4me2 and the shortened HMD then protects a region from H3K27me3 accumulation, resulting in consolidation of the long-term expressions of developmental genes. Consistently, we have previously demonstrated two distinct regulations of *zic1/zic4* transcription: the expression in somites is induced by secreted signals but is later maintained throughout life in a cell-autonomous manner (Kawanishi et al., 2013). Similar to *zic1/zic4*, it has been reported that the embryonic Hox expression pattern is epigenetically maintained in fibroblasts of the human adult foot and is required to maintain its site-specific identity (Rinn et al., 2008). Notably in our dataset, very large K27HMDs covering Hox clusters observed at the blastula stage undergo HMD shortening at active loci in the adult myotome. Given that DNA methylation is a stable epigenetic modification (Smith and Meissner, 2013), this K27HMD shortening could serve as cellular memory by marking developmental genes once activated. We propose the two-step regulation of key developmental genes: histone modification-dependent induction and HMD shortening-dependent long-term maintenance.

Our comparative analysis demonstrated that the overall pattern of DNA methylation is conserved between medaka and human. Moreover, the majority of gene promoters shared the same epigenetic states between the two systems. In particular, large K27HMDs were highly conserved; thus, pre-marking of key transcription factor genes by a large K27HMD and its strong

Development (2014) 141, 2568-2580 doi:10.1242/dev.108548

repression are shared features in the vertebrate development. Importantly, our analysis also identified human-specific K27HMD genes and found that those methylated in medaka pluripotent cells (Class II) are related to neuronal activities. Human-specific K27HMDs had lower sequence conservation than those common to both human and medaka. Indeed, regulatory elements under human constraint, but not conserved in mammals, were found to be associated with neuronal activities (Ward and Kellis, 2012). These neuronal genes are not expressed in both medaka and human pluripotent cells, but the repression is mediated by distinct modifications: DNA methylation and H3K27me3, respectively. What, then, is the biological significance of such differential epigenetic marking? One possibility is that K27HMD may enable flexible changes in regulation of those genes; the poised state or sustained expression can be achieved through regulation of histone modification and of HMD size. Interestingly, a significant number of glutamate receptor genes were found at Class II human-specific K27HMDs. Glutamate receptors are known to have a key function in synaptic plasticity and are important for learning and memory (Lancaster and Dalmau, 2012; Salter and Kalia, 2004). The flexible regulation of these genes could be required for the sophisticated neuronal network in human.

MATERIALS AND METHODS

Fish strains

We used medaka d-rR strains as wild type. The *Da* mutant used in this study was the same strain as previously described (Moriyama et al., 2012). We crossed the *Da* homozygous mutant and d-rR to generate *Da* heterozygous mutants. Medaka fish were maintained and raised under standard conditions. Identification of developmental stages was performed as previously described (Iwamatsu, 2004).

Whole-mount *in situ* hybridization

Whole-mount *in situ* hybridization was performed as previously described (Takashima et al., 2007). For *tbx2* (ENSORLG00000014792 and ENSORLG00000010011) primers for DIG-labeled RNA probes are listed in supplementary material Table S10. For *pax6* (ENSORLG00000009913 and ENSORLG00000000847) 3′UTR was cloned by SMARTer RACE cDNA Amplification kit (Clontech) and used for RNA probe synthesis. Primers are listed in supplementary material Table S10.

RT-qPCR

RNA was isolated from adult wild-type and *Da* mutant muscle (dorsal and ventral parts were separated) using ISOGEN (Nippon Gene), according to the manufacturer's protocol. SuperSucript III (Invitrogen) was used for cDNA synthesis. qPCR was performed with the Stratagene Mx3000P system (Agilent Technologies) using the THUNDERBIRD SYBR qPCR mix (Toyobo). All primers are listed in supplementary material Table S10.

Bisulfite sequence

Genomic DNA was isolated from adult and larva muscle (dorsal and ventral parts were separated) and bisulfite treatment was performed using methyl easy DNA Bisulphite Modification Kit (Human Genetic Signatures). Bisulfite-converted DNA was subjected to PCR using Ex Taq (TaKaRa) and TOPO-TA cloning (Life Technologies). Amplified fragments were sequenced and analyzed and visualized using the QUMA software (Kumaki et al., 2008). All primers are listed in supplementary material Table S10.

Whole-genome bisulfite sequence (WGBS)

For adult and larva muscle (dorsal and ventral parts were separated), sample preparation, library construction, sequencing and mapping for WGBS were performed as previously described (Qu et al., 2012). Briefly, genomic DNA from medaka was isolated and sonicated to a desired size range (100-400 bp). The DNA fragments were treated with DNA polymerase to generate blunt ends and were ligated with double-stranded DNA adaptors containing methylated cytosines, which were designed to amplify only those DNA fragments carrying bisulfite-converted adaptor sequences at both ends. Followed by 7-10 cycles of PCR, 250-450 bp fractionated DNA was sequenced using an Illumina GAIIx genome analyzer. We converted all cytosines in reads and in both the Watson and Crick strands of the reference genome to thymines for primary mapping, and used Smith-Waterman alignments between the original sequences of primary best hits. The level of methylation of a particular cytosine was estimated by dividing the number of mapped reads reporting a cytosine (C) by the total number of reads reporting a C or T (thymine).

ChIP

ChIP was performed using the following antibodies: H3K27me3 (Millipore, 07-449), H3K4me1 (Millipore, 07-436), H3K4me2 (Millipore, 07-030), H3K4me3 (Millipore, 07-473) and H3K27Ac (abcam, ab4729). ChIP was peformed as previously described with modifications (Lindeman et al., 2009). Cells were dissociated using a 21G needle and fixed with 1% formaldehyde for 10 min at RT then quenched by adding glycine (200 mM final). After washing with PBS containing 20 mM Na-butylate, complete protease inhibitor (Roche) and 1 mM PMSF, cell pellets were suspended in lysis buffer [50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 1% SDS, 20 mM Na-butylate, complete protease inhibitors, 1 mM PMSF], sonicated ten times using sonifier (Branson) at power five, and centrifuged for collecting chromatin lysates. The chromatin lysates were diluted with ChIP RIPA buffer [10 mM Tris-HCl (pH 8.0), 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton-X100, 0.1% SDS, 0.2% sodium deoxycholate, 20 mM Na-butylate, complete protease inhibitors, 1 mM PMSF] and rotated with an antibody/protein A Dynabeads complex overnight at 4°C. Immunoprecipitated materials were washed three times with ChIP RIPA buffer and once with TE buffer, followed by elution with Lysis buffer at 65°C over night. Eluted samples were treated with RNase A for 2 h at 37°C and proteinase K for 2 h at 55°C, and DNA was purified by phenol/chloroform extraction and ethanol precipitation. Input DNA was simultaneously treated from the elution step. Input and immunoprecipitated DNAs were analyzed by quantitative PCR with the Stratagene Mx3000P system (Agilent Technologies) using the THUNDERBIRD SYBR qPCR mix (TOYOBO). All primers are listed in supplementary material Table S10.

ChIP-seq

For ChIP-seq analysis, we used ~10^6 cells for one modification. Using the input and immunoprecipitated samples of ChIP, ChIP-seq templates were prepared using the TruSeq DNA sample prep kit (version 2; Illumina) and sequenced using Genome analyzer IIx (Illumina). The sequence was read by 36-base single-end sequencing.

RNA-seq

RNA was isolated from adult muscle (dorsal and ventral parts were separated) and adult liver using RNeasy mini kit (Qiagen) or ISOGEN (Nippon Gene), according to the manufacturer's protocol. Purified RNAs were treated with Ribominus eukaryote kit for RNA-seq (Life Technologies). RNA-seq analysis was conducted basically following the instructions from the manufacturer. Briefly, the RNA-seq template was prepared using TruSeq RNA-seq sample prep kit (version 2; Illumina), omitting the polyA selection procedure. The double-stranded PCR products were purified and size fractionated using a bead-mediated method (AMPure, Ambion). Sequencing was conducted on Genome analyzer IIx platform (Illumina) using TruSeq Cluster generation kit. At least 20 million sequences of a 36-base single-reads were generated per sample.

Data access

All sequence data are deposited at the NCBI Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra) (accession number SRP029233).

Data analyses

The methods for data analyses are described in detail in the supplementary material.

Acknowledgements
We thank A. Terashima for critical reading of the manuscript. We are grateful to Y. Ozawa for fish care.

Competing interests
The authors declare no competing financial interests.

Author contributions
R.N., H.T. and S.M. designed the study. R.N. performed the experiments, analyzed data and drafted the manuscript. H.T., S.M., W.Q. and T.T. supervised the research and carried out revisions of the manuscript for important intellectual content. T.O. performed ChIP using medaka liver. K.O., K.M. and S.H. constructed the library for Bisulfite-seq. S.S. and Y.S. conducted sequencing of Bisulfite-seq, ChIP-seq and RNA-seq. W.Q. performed mapping of sequencing data to the medaka genome. T.L.S. conducted the visualization of sequencing data using UTGB. K.I. performed ChromHMM analysis.

Funding
This research was supported in part by Grants-in-Aid for Scientific Research Priority Area Genome Science and Scientific Research on Innovative Areas 'Logic of Biological Pattern Formation' [22127002], Global Center of Excellence Program 'Integrative Life Science' from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT); by grants from the Uehara Memorial Foundation and the Mitsubishi Foundation to H.T.; by Grant-in-Aid for Scientific Research (A) [23241058], Innovative Cell Biology by Innovative Technology, Global COE program 'Deciphering Biosphere from Genome Big Bang' from MEXT to S.M.; and by Grant-in-Aid for Young Scientists (B) [24770202 from MEXT to T.T.]. Sequencing for genome wide analyses was in part supported by MEXT KAKENHI [221S0002]. R.N. is a recipient of the Fellowship of the Japan Society for the Promotion of Science for Junior Scientists.

Supplementary material
Supplementary material available online at
http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.108548/-/DC1

References
Aizawa, K., Shimada, A., Naruse, K., Mitani, H. and Shima, A. (2003). The medaka midblastula transition as revealed by the expression of the paternal genome. *Gene Expr. Patterns* **3**, 43-47.
Akkers, R. C., van Heeringen, S. J., Jacobi, U. G., Janssen-Megens, E. M., Françoijs, K.-J., Stunnenberg, H. G. and Veenstra, G. J. (2009). A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos. *Dev. Cell* **17**, 425-434.
Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K. et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326.
Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6-21.
Bogdanovic, O., Long, S. W., van Heeringen, S. J., Brinkman, A. B., Gomez-Skarmeta, J. L., Stunnenberg, H. G., Jones, P. L. and Veenstra, G. J. C. (2011). Temporal uncoupling of the DNA methylome and transcriptional repression during embryogenesis. *Genome Res.* **21**, 1313-1327.
Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K. et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-353.
Brinkman, A. B., Gu, H., Bartels, S. J. J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A. et al. (2012). Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* **22**, 1128-1138.
Cedar, H. and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* **10**, 295-304.
Darnell, J. E., Jr. (2002). Transcription factors as targets for cancer therapy. *Nat. Rev. Cancer* **2**, 740-749.
Deaton, A. M. and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010-1022.
ENCODE Project Consortium (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9**, e1001046.
Ernst, J. and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215-216.
Fu, Y., Sinha, M., Peterson, C. L. and Weng, Z. (2008). The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.* **4**, e1000138.
Fujikura, J., Yamato, E., Yonemura, S., Hosoda, K., Masui, S., Nakao, K., Miyazaki, J.-i. and Niwa, H. (2002). Differentiation of embryonic stem cells is induced by GATA factors. *Genes Dev.* **16**, 784-789.
Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. and Noble, W. S. (2007). Quantifying similarity between motifs. *Genome Biol.* **8**, R24.
Hagarman, J. A., Motley, M. P., Kristjansdottir, K. and Soloway, P. D. (2013). Coordinate regulation of DNA methylation and H3K27me3 in mouse embryonic stem cells. *PLoS ONE* **8**, e53880.
Handoko, L., Xu, H., Li, G., Ngan, C. Y., Chew, E., Schnapp, M., Lee, C. W. H., Ye, C., Ping, J. L. H., Mulawadi, F. et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.* **43**, 630-638.
Harrelson, Z., Kelly, R. G., Goldin, S. N., Gibson-Brown, J. J., Bollag, R. J., Silver, L. M. and Papaioannou, V. E. (2004). Tbx2 is essential for patterning the atrioventricular canal and for morphogenesis of the outflow tract during heart development. *Development* **131**, 5041-5052.
Hendrich, B. and Tweedie, S. (2003). The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet.* **19**, 269-277.
Hu, J.-L., Zhou, B. O., Zhang, R.-R., Zhang, K.-L., Zhou, J.-Q. and Xu, G.-L. (2009). The N-terminus of histone H3 is required for de novo DNA methylation in chromatin. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22187-22192.
Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44-57.
Iwamatsu, T. (2004). Stages of normal development in the medaka Oryzias latipes. *Mech. Dev.* **121**, 605-618.
Jaillon, O., Aury, J.-M., Brunet, F., Petit, J.-L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A. et al. (2004). Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature* **431**, 946-957.
Jeong, M., Sun, D., Luo, M., Huang, Y., Challen, G. A., Rodriguez, B., Zhang, X., Chavez, L., Wang, H., Hannah, R. et al. (2014). Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat. Genet.* **46**, 17-23.
Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., Kasai, Y. et al. (2007). The medaka draft genome and insights into vertebrate genome evolution. *Nature* **447**, 714-719.
Kawanishi, T., Kaneko, T., Moriyama, Y., Kinoshita, M., Yokoi, H., Suzuki, T., Shimada, A. and Takeda, H. (2013). Modular development of the teleost trunk along the dorsoventral axis and zic1/zic4 as selector genes in the dorsal module. *Development* **140**, 1486-1496.
Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S. et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* **4**, e1000242.
Kumaki, Y., Oda, M. and Okano, M. (2008). QUMA: quantification tool for methylation analysis. *Nucleic Acids Res.* **36**, W170-W175.
Lancaster, E. and Dalmau, J. (2012). Neuronal autoantigens–pathogenesis, associated disorders and antibody testing. *Nat. Rev. Neurol.* **8**, 380-390.
Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C. T., Low, H. M., Kin Sung, K. W., Rigoutsos, I., Loring, J. et al. (2010). Dynamic changes in the human methylome during differentiation. *Genome Res.* **20**, 320-331.
Lindeman, L. C., Vogt-Kielland, L. T., Aleström, P. and Collas, P. (2009). Fish'n ChIPs: chromatin immunoprecipitation in the zebrafish embryo. *Methods Mol. Biol.* **567**, 75-86.
Lindeman, L. C., Andersen, I. S., Reiner, A. H., Li, N., Aanes, H., Østrup, O., Winata, C., Mathavan, S., Müller, F., Alestrom, P. et al. (2011). Prepatterning of developmental gene expression by modified histones before zygotic genome activation. *Dev. Cell* **21**, 993-1004.
Lindroth, A. M., Park, Y. J., McLean, C. M., Dokshin, G. A., Persson, J. M., Herman, H., Pasini, D., Miró, X., Donohoe, M. E., Lee, J. T. et al. (2008). Antagonism between DNA and H3K27 methylation at the imprinted Rasgrf1 locus. *PLoS Genet.* **4**, e1000145.
Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M. et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322.
Lister, R., Pelizzola, M., Kida, Y. S., Hawkins, R. D., Nery, J. R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S. et al. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471**, 68-73.
Long, H. K., Sims, D., Heger, A., Blackledge, N. P., Kutter, C., Wright, M. L., Grutzner, F., Odom, D. T., Patient, R., Ponting, C. P. et al. (2013). Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *Elife* **2**, e00348.
Merzdorf, C. S. (2007). Emerging roles for zic genes in early development. *Dev. Dyn.* **236**, 922-940.
Mohn, F., Weber, M., Rebhan, M., Roloff, T. C., Richter, J., Stadler, M. B., Bibel, M. and Schübeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **30**, 755-766.
Molaro, A., Hodges, E., Fang, F., Song, Q., McCombie, W. R., Hannon, G. J. and Smith, A. D. (2011). Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* **146**, 1029-1041.
Moriyama, Y., Kawanishi, T., Nakamura, R., Tsukahara, T., Sumiyama, K., Suster, M. L., Kawakami, K., Toyoda, A., Fujiyama, A., Yasuoka, Y. et al.

(2012). The medaka zic1/zic4 mutant provides molecular insights into teleost caudal fin evolution. *Curr. Biol.* **22**, 601-607.

Ooi, S. K. T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C. D. et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714-717.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G. A., Stewart, R. and Thomson, J. A. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* **1**, 299-312.

Potok, M. E., Nix, D. A., Parnell, T. J. and Cairns, B. R. (2013). Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. *Cell* **153**, 759-772.

Puschel, A. W., Gruss, P. and Westerfield, M. (1992). Sequence and expression pattern of pax-6 are highly conserved between zebrafish and mice. *Development* **114**, 643-651.

Qu, W., Hashimoto, S.-i., Shimada, A., Nakatani, Y., Ichikawa, K., Saito, T. L., Ogoshi, K., Matsushima, K., Suzuki, Y., Sugano, S. et al. (2012). Genome-wide genetic variations are highly correlated with proximal DNA methylation patterns. *Genome Res.* **22**, 1419-1425.

Rinn, J. L., Wang, J. K., Allen, N., Brugmann, S. A., Mikels, A. J., Liu, H., Ridky, T. W., Stadler, H. S., Nusse, R., Helms, J. A. et al. (2008). A dermal HOX transcriptional program regulates site-specific epidermal fate. *Genes Dev.* **22**, 303-307.

Rush, M., Appanah, R., Lee, S., Lam, L. L., Goyal, P. and Lorincz, M. C. (2009). Targeting of EZH2 to a defined genomic site is sufficient for recruitment of Dnmt3a but not de novo DNA methylation. *Epigenetics* **4**, 404-414.

Salter, M. W. and Kalia, L. V. (2004). Src kinases: a hub for NMDA receptor regulation. *Nat. Rev. Neurosci.* **5**, 317-328.

Sasaki, S., Mello, C. C., Shimada, A., Nakatani, Y., Hashimoto, S.-i., Ogawa, M., Matsushima, K., Gu, S. G., Kasahara, M., Ahsan, B. et al. (2009). Chromatin-associated periodicity in genetic variation downstream of transcriptional start sites. *Science* **323**, 401-404.

Segal, E. and Widom, J. (2009). What controls nucleosome positions? *Trends Genet.* **25**, 335-343.

Smith, Z. D. and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204-220.

Suzuki, M. M. and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* **9**, 465-476.

Takashima, S., Shimada, A., Kobayashi, D., Yokoi, H., Narita, T., Jindo, T., Kage, T., Kitagawa, T., Kimura, T., Sekimizu, K. et al. (2007). Phenotypic analysis of a novel chordin mutant in medaka. *Dev. Dyn.* **236**, 2298-2310.

Taylor, J. S., Braasch, I., Frickey, T., Meyer, A. and Van de Peer, Y. (2003). Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Res.* **13**, 382-390.

Tweedie, S., Charlton, J., Clark, V. and Bird, A. (1997). Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Mol. Cell. Biol.* **17**, 1469-1475.

Valouev, A., Johnson, D. S., Sundquist, A., Medina, C., Anton, E., Batzoglou, S., Myers, R. M. and Sidow, A. (2008). Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat. Methods* **5**, 829-834.

Valouev, A., Johnson, S. M., Boyd, S. D., Smith, C. L., Fire, A. Z. and Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. *Nature* **474**, 516-520.

Vastenhouw, N. L., Zhang, Y., Woods, I. G., Imam, F., Regev, A., Liu, X. S., Rinn, J. and Schier, A. F. (2010). Chromatin signature of embryonic pluripotency is established during genome activation. *Nature* **464**, 922-926.

Ward, L. D. and Kellis, M. (2012). Evidence of abundant purifying selection in humans for recently acquired regulatory functions. *Science* **337**, 1675-1678.

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I. and Young, R. A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319.

Woo, C. J., Kharchenko, P. V., Daheron, L., Park, P. J. and Kingston, R. E. (2010). A region of the human HOXD cluster that confers polycomb-group responsiveness. *Cell* **140**, 99-110.

Wu, H., Coskun, V., Tao, J., Xie, W., Ge, W., Yoshikawa, K., Li, E., Zhang, Y. and Sun, Y. E. (2010). Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science* **329**, 444-448.

Wu, S.-F., Zhang, H. and Cairns, B. R. (2011). Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm. *Genome Res.* **21**, 578-589.

Xie, W., Schultz, M. D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J. W., Tian, S., Hawkins, R. D., Leung, D. et al. (2013). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134-1148.

Yi, M., Hong, N. and Hong, Y. (2009). Generation of medaka fish haploid embryonic stem cells. *Science* **326**, 430-433.

Zhao, X. D., Han, X., Chew, J. L., Liu, J., Chiu, K. P., Choo, A., Orlov, Y. L., Sung, W.-K., Shahab, A., Kuznetsov, V. A. et al. (2007). Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* **1**, 286-298.

Zhou, V. W., Goren, A. and Bernstein, B. E. (2011). Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* **12**, 7-18.