Title

Lhx2 regulates temporal changes in chromatin accessibility and transcription factor binding in retinal progenitor cells.

Authors

Cristina Zibetti[1], Sheng Liu[2,6], Jun Wan[2,6], Jiang Qian[2*], Seth Blackshaw[1-5] *.

Affiliations

[1]Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA. [2]Department of Ophthalmology, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA. [3]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA. [4]Center for Human Systems Biology, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA. [5]Institute for Cell Engineering, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA, [6]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis IN 46202.

*Correspondence to sblack@jhmi.edu; jiang.qian@jhmi.edu

Abstract: Retinal progenitor cells (RPCs) pass through multiple stages of developmental competence, where they successively acquire and lose the ability to generate individual cell subtypes. To identify the transcriptional regulatory networks that control these transitions, we conducted epigenomic and transcriptomic profiling of early and late-stage RPCs and observed a developmentally dynamic landscape of chromatin accessibility. Open chromatin regions that showed stage-specificity, as well as those shared by early and late-stage RPCs, were selectively targeted by the homeodomain factor Lhx2, which is expressed throughout retinal neurogenesis but also regulates many stage-specific processes in RPCs. Stage-specific Lhx2 binding sites were frequently associated with target sites for transcription factors that are preferentially expressed in either early or late-stage RPCs, and which were predicted to possess pioneer activity. *Lhx2* loss of function in RPCs led to a loss of chromatin accessibility at both direct Lhx2 target sites and more broadly across the genome, as well as a loss of binding by transcription factors associated with stage-specific Lhx2 target sites. These findings demonstrate a central role for Lhx2 in control of chromatin accessibility in RPCs, and identify transcription factors that may guide stage-specific target site selection by Lhx2.

Summary: Lhx2 is a central regulator of chromatin accessibility in retinal progenitor cells, and interacts with stage-specific transcription factors to regulate genes that are dynamically expressed during retinal neurogenesis.

MAIN TEXT

Introduction:

During the course of neurogenesis, neural progenitors move through different states of developmental competence, in which they successively acquire and lose their ability to generate individual cell subtypes. This process, also known as temporal patterning, occurs throughout the nervous system in both vertebrates and invertebrates (*1, 2*). These transitions are largely cell-autonomous and proceed in a unidirectional and irreversible manner, and are driven by changes in transcription factor (TF) expression. The murine retina is an excellent model for mechanistic analysis of temporal patterning, as the generation of all major cell types is controlled through this process. A limited number of TFs have been identified that show differential expression in early and late-stage retinal progenitor cells (RPCs), and which also control changes in developmental competence (*3-6*). Other TFs expressed throughout retinal neurogenesis, such as the LIM homeodomain factor *Lhx2*, are nonetheless essential for regulating multiple stages of RPC competence. Lhx2 is required for the transition of early-stage RPCs from an immature state, in which they can generate retinal ganglion cells (RGCs), to a more mature state in which they can no longer do so (*7*). In contrast, in late-stage RPCs, Lhx2 is essential for generation of late-born Muller glia and does this in part by enhancing Notch signaling (*8, 9*). Lhx2 regulates these temporally dynamic processes in parallel with transcriptional programs common to early and late-stage RPCs, where it maintains proliferative and neurogenic competence, and represses expression of genes that are specific to the anterodorsal hypothalamus and thalamic eminence (*10*).

The mechanism by which these highly stage-specific TFs interact with broadly expressed factors such as Lhx2 to regulate retinal cell fate specification is unknown, and a broader picture of the transcriptional regulatory networks that control RPC competence is likewise lacking. In this study, in order to identify *cis* and trans-acting factors that regulate temporal patterning in the developing retina, we globally profiled chromatin accessibility and mRNA expression in both early and late-stage RPCs, using ATAC-Seq and RNA-Seq, and identify Lhx2 as a central regulator of changes in chromatin accessibility that occur in RPCs during the course of neurogenesis.

Results:

Lhx2 binding sites are highly enriched in open chromatin in RPCs

Fluorescence-activated cell sorting (FACS) was used to isolate RPCs from mice expressing the RPC-specific *Chx10-Cre:eGFP* transgene (*11*) from embryonic day (E)14 and postnatal day (P)2 retina, representing early and late stages of RPC competence, respectively (*1*). The great majority of GFP-positive cells expressed the RPC-specific markers Chx10, Ki67 and Ccnd1, while these were absent from the GFP-negative postmitotic fraction (Fig. S1A-G). Flow-sorted cells were profiled by RNA-Seq and ATAC-Seq, relying on direct *in vitro* transposition of sequencing adapters into native chromatin (*12*). ChIP-Seq was then performed on a select transcription factor candidate, identified in the preliminary screening of the ATAC-Seq derived open chromatin regions. ChIP-Seq binding sites for the candidate factor were ultimately integrated and compared with the age-matched RNA-Seq and ATAC-Seq profiles from control and conditional

Lhx2 knockout retinas (Fig.1A, S1H, I). ATAC-Seq data was highly reproducible among experimental replicates (Fig. S2A-D), and identified open chromatin regions (OCRs) that were shared between time points, or were specific to either E14 or P2 (Fig. 1B). These represent 2.4%, 0.72%, and 0.80% of the genome, respectively.

We next scanned OCRs common to E14 and P2 RPCs for candidate TF binding sites. Clustering of the assigned position weight matrices (PWM) revealed that homeobox-containing TFs constitute the prevalent cluster, representing 284/373 putative matches (Fig. 1C, highlighted in the inset), followed by C2H2-type, GC-rich and KLF zinc finger proteins; followed by POU-homeodomain factors, MADS box and CCAAT-binding factors. The consensus motif for the insulator protein CTCF, involved in controlling chromatin looping, is also highly enriched in OCRs. To pinpoint the individual TFs that showed strong enrichment in common OCRs, we examined the gene expression level of the TFs in RPCs. Interestingly, only a few of the TFs that were the top predicted matches for the homeobox-containing factors were detectably expressed in RPCs by RNA-Seq (Fig. 1D), and many represented close variants of the Lhx2 consensus sequence (Fig. 1E, S2E). Indeed, among the TFs that are expressed in RPCs, Lhx2 consensus site was identified as the most represented TF motif in OCRs common to early and late RPCs (Fig. 1F, Auxiliary table 1), and variably represented in stage-specific OCRs (Auxiliary table 2,3). Lhx2 occupancy resulted into detectable footprints at both stages (Fig. 1G). Finally, ChIP-Seq was performed for Lhx2 in E14 and P2 mouse retina. Integration with age-matched ATAC-Seq profiles from flow-sorted RPCs demonstrated that Lhx2 ChIP-Seq peaks do indeed overlap with many of OCRs (Fig.1H, 2A, S3A-G).

Lhx2 target site selection in RPCs is dynamic during the course of retinal neurogenesis

Lhx2 ChIP-Seq peaks are predominantly located in intragenic and intronic regions of the genome (Fig. S4A), are enriched in promoter regions, 5'UTR, exons, and ncRNA at both time points (Fig. 2B), and are enriched for functions related to control of retinal development (Fig. S4B,C). The Lhx2 ChIP-Seq data (Fig. S3G) was then compared to a manually curated list of genes that showed enriched expression in individual retinal cell types, which also included RNA-Seq data from E14 and P2 RPCs (Auxiliary table 4,5). Lhx2 ChIP-Seq peaks were preferentially associated with genes that were selectively expressed in late-stage RPCs, Müller glia, and the embryonic anterodorsal hypothalamus at both E14 and P2 (Fig. 2C, Auxiliary table S5). Peaks found only at E14 were also selectively associated with genes expressed in early-stage RPCs (Fig. 2D), while genes enriched in late-stage RPCs were overrepresented in the P2 ChIP-Seq dataset (Fig. 2D).

To identify genes whose expression is directly regulated by Lhx2 in early and late-stage RPCs, Lhx2 targets identified by ChIP-Seq were compared to the RNA-Seq data from the E14 and P2 GFP-positive fractions and to the expression profiles obtained from *Lhx2*-deficient cells isolated at E14 (*Chx10-Cre;Lhx2*$^{lox/lox}$ retinas) (Fig. 2D,E, Fig. S5A-C) and P2 (*Lhx2*$^{lox/lox}$ RPCs electroporated with pCAG-Cre-GFP at P0 and isolated by FACS at P2) (Fig. 2D,F, Fig. S5D-F). In total, more than 30% of the Lhx2 peaks shared between time points and 20% of the stage-specific peaks were associated with differentially expressed genes (DEG) detected from the comparison between E14 and P2 RPCs, potentially accounting for 10 to 14% of the differentially expressed transcriptome.

Most of the genes that were differentially expressed between early and late RPCs, as well as those that were enriched in early and late RPCs compared to age-matched GFP-negative postmitotic cells, were associated with stage-specific Lhx2 peaks (Auxiliary table 4). Importantly, 47% of all genes that were differentially expressed between E14 and P2 RPCs also showed altered expression in *Lhx2* cKO samples, at one or both of these timepoints (Fig. S5G, Auxiliary table 6,7). Examples of such temporally dynamic and *Lhx2*-regulated genes include *Ndnf*, which is upregulated in P2 *Lhx2* cKO cells, and the late-stage RPC and MG-enriched *Car2* locus, which is downregulated in P2 *Lhx2* cKO cells (Fig. S5H). These results show that Lhx2 plays a central role in controlling temporally dynamic gene expression in RPCs.

Lhx2 selectively targets OCRs and regulates genes associated with RPC function

Lhx2 binding sites are predominantly associated with open chromatin. In E14 RPCs, 4.3% of OCRs overlapped with Lhx2 ChIP-Seq peaks and 3.5% of P2 OCRs were associated with age-matched Lhx2 peaks, with a higher fraction of closely co-localizing peaks at E14 compared to P2 (Auxiliary table 6). The majority of Lhx2 target sites fall in OCRs: 76.7% of the E14 Lhx2 *cis*-regulatory sites fall in OCRs in E14 GFP-positive RPCs, while at P2 61.7% of the P2 ChIP-Seq peaks co-occur with P2 GFP-positive OCRs (Fig. S4D, Auxiliary table 7). Notably, Lhx2 ChIP-Seq peaks were preferentially co-localized with positioned nucleosomes at E14 (Fig. 2G) but bimodal at P2 (Fig. 2H) suggesting Lhx2 may compete for nucleosome binding in early-stage RPCs (Fig. 2G).

Overall, 9.9% of the nucleosome-interspersed OCRs identified in E14 RPCs, and 9.3% of those identified at P2, are associated with bidirectionally-transcribed Lhx2 binding sites. These represent 19.3% of the high confidence Lhx2 ChIP-Seq peaks at E14 and 17.4% at P2. Lhx2 targets that flanked positioned nucleosomes within OCRs were enriched at 5'UTR and promoter regions at both E14 (Fig. 2G) and P2 (Fig. 2H) and coupled with bimodally distributed, likely active enhancers, marked by H3K27ac peaks at P2. Since exons and ncRNAs were also found among the enriched categories of Lhx2 target sites at both ages, it is likely that a subset of these *cis*-regulatory regions may encode enhancer-associated RNAs (eRNAs).

Lhx2 targets in E14 RPCs were associated with genes involved in regulatory RNA pathways, promoter opening and cell cycle checkpoints (Fig. 2I) and those in P2 RPCs population were enriched for functional categories related cell cycle progression, Notch signaling and DNA replication, as well as phenotypes such as retina hypoplasia (Fig. 2J). Enrichment for these functional categories was confirmed by comparison with *Lhx2*-dependent gene sets identified by RNA-Seq analysis of the age-matched *Lhx2* cKO samples.

Identification of TF motifs associated with Lhx2 ChIP-Seq peaks

We next investigated which TFs could co-localize with Lhx2 ChIP-Seq peaks at E14 and P2, or at each individual timepoint. The Lhx2 consensus sequence and closely related sequences were over-represented in peaks present at both E14 and P2 (Fig. 3A,B). Multiple motifs were found within a broad range of similarity to the known Lhx2 consensus (Pearson's correlation > 0.6), possibly underlying differences in affinity and/or combinatorial interaction with other transcription factors (Fig. S6A-I).

182 In the E14 retina, the SoxB1 related family members (Sox2) *(15)* and MADS
183 transcription factors were found at E14 (Fig. 3A) while Kruppel like factors (Klf9/13)
184 *(13),* NF-I family members (Nfia/Nfib/Nfic/Nfix) *(16)*, SoxE (Sox8/9) *(17)* and Ascl1
185 *(18)* were found at P2 (Fig.3B). When only known TF motifs are considered, (Fig.
186 S7A,B), preferential representation of homeobox/bHLH transcription factors was
187 observed at E14 (Fig. 3C) and NF-I, CTCF and Forkhead family members observed at P2
188 (Fig. 3D).
189
190 **Lhx2 broadly regulates chromatin accessibility in RPCs**
191 We then tested whether loss of function of Lhx2 affected chromatin accessibility
192 in RPCs. We first analyzed ATAC-Seq profiles from control samples at E14 (whole
193 retina) and P2 (FACS-isolated cells electroporated at P0) (Fig. S8A-G), and observed
194 high overall correlation with ATAC-Seq profiles from age-matched purified RPCs. By
195 comparing these to ATAC-Seq profiles from E14 and P2 *Lhx2* cKO samples, we
196 observed an extensive reduction in OCRs (Fig. S8F,G).
197 As expected, we observed a substantial reduction in Lhx2 footprints in both E14
198 and P2 *Lhx2* cKO samples. The average read coverage was reduced 4-fold in E14 *Lhx2*
199 cKO retinas (Fig. 4A), while in P2 *Lhx2* cKO retinas a 1.5-fold reduction in the average
200 read coverage was observed (Fig. 4B). The reduction in footprint inflection depth was
201 even more pronounced, with an 8-fold reduction observed at E14 and a 1.6-fold reduction
202 at P2. Finally, 96% of all Lhx2 footprints identified in ATAC-Seq profiles of E14 RPCs
203 were lost in E14 *Lhx2* cKO retinas, while 39% were lost at P2 (Auxiliary table 8), and
204 mean footprint signal was significantly reduced at both ages (Auxiliary table 9). This
205 difference between E14 and P2 may partially reflect perdurance of Lhx2 expression at P2
206 following Cre-mediated deletion from P0 onward. Local accessibility of Lhx2 target
207 sites correlated with Lhx2 occupancy, both at E14 (Fig. S8H) and P2 (Fig. S8I), and
208 significant reductions were observed in *Lhx2* cKO retinas at both ages. Because local
209 accessibility of Lhx2 targets was reduced at both E14 and P2 (Fig. 4C,D upper panel, Fig.
210 S8F,G; Auxiliary table 10), we hypothesized that Lhx2 might play a central role in
211 maintenance of Lhx2-targeted OCRs in RPCs. 40% of all OCRs associated with Lhx2
212 ChIP-Seq peaks are lost in E14 *Lhx2* cKO samples, while 8% are lost at P2 (Auxiliary
213 table 10). The effect of *Lhx2* loss of function on chromatin accessibility was not merely
214 focal, but extended genome-wide. Overall, 50% of OCRs in the E14 mutants and 22% in
215 P2 mutants were lost (Fig. 4C,D lower panel, Auxiliary table 11), suggesting that Lhx2 is
216 broadly required in RPCs to maintain accessibility of these regions.
217
218 **Lhx2 affects local and global chromatin accessibility in RPCs by regulating**
219 **expression and DNA binding of pioneer TFs**
220 This finding raised the possibility that loss of function of *Lhx2* may disrupt the
221 transcription and/or function of TFs with pioneering activity that are expressed in RPCs,
222 and that these TFs may in turn regulate accessibility of these OCRs. Consistent with this
223 hypothesis, multiple RPC-expressed TFs with predicted pioneering function showed
224 reduced expression and/or reduced accessibility in nearby OCRs in *Lhx2* cKO samples
225 (Fig.4E,F, Auxiliary table 7). Interestingly, a small number of such TFs, such as Dlx2
226 and Atoh8, showed both reduced accessibility at nearby OCRs and upregulated

227 expression in P2 *Lhx2* cKO samples, implying that these OCRs might mediate Lhx2-
228 depedent transcriptional repression of these genes (Fig. 4F, Auxiliary table 12).
229
230 To identify RPC-expressed TFs that might target developmentally dynamic OCRs
231 in parallel with Lhx2, we next analyzed the sequence of OCRs that were selectively
232 active at either E14 or P2. We identified multiple TFs predicted to target RPC-associated
233 OCRs. Some had temporally dynamic expression in RPCs, had known or putative pioneer
234 function (Auxiliary table 13) and were predicted to target Lhx2 ChIP-Seq peaks (Fig.
235 3C,D)(*12*).
236
237 Overall, TFs that showed enriched expression in E14 or P2 RPCs showed higher
238 predicted pioneering function than TFs that showed similar expression at both ages (Fig.
239 S9). TFs with known or predicted pioneering activity that are enriched in E14 RPCs
240 include *Meis1*. In P2 RPCs, these include members of several TF subfamilies, including
241 SoxE (*Sox8/9*), NFI (*Nfia/Nfib/Nfix*) (*13*), *Ascl1*, *Hes5*, and KLF (*Klf9/13*) (*14*). Several
242 of these TFs control development of late-born cell types (*8, 15-17*). However, other TFs
243 that showed equivalent levels of mRNA expression in E14 and P2 RPCs – such as
244 Foxm1, Olig2, and Pax6 – preferentially targeted OCRs that were selectively active in
245 either early or late-stage RPCs (Fig. 3C,D; Auxiliary table 8). Multiple TFs expressed in
246 RPCs – such as *Sox2* (*18*), Smad3, ZFX (Zfhx4) and Etv5 – that exhibit a high pioneer
247 potential were directly transcriptionally targeted by Lhx2 (Auxiliary table 14). These
248 TFs also showed reductions in both footprint count and mean footprint signal (Fig. 4G,
249 Auxiliary table 8, 9) following loss of *Lhx2* function.
250
251 **Dynamic Lhx2-dependent regulation of DNA binding by RPC-expressed TFs**
252 **with predicted pioneer activity**
253 We next conducted a more detailed analysis of our ATAC-Seq data to investigate
254 how loss of function of *Lhx2* at both E14 and P2 affected binding by selected candidate
255 pioneer TFs. As expected, the average insertion frequency at Lhx2 motifs was
256 substantially reduced in both E14 and P2 *Lhx2*-deficient retinas (Fig. 5A). In line with the
257 mRNA expression patterns of their associated TFs, NF-I and KLF footprints are
258 overrepresented in P2 compared to E14 OCRs, while Sox2 showed comparable
259 occupancy in early and late RPCs (Auxiliary table 8,9). The total number of footprints
260 and the mean signal associated with footprints of gliogenic factors, such as NF-I
261 (Nfia/b/x) were reduced in *Lhx2* cKO, consistent with the disruption of gliogenesis that
262 occurs following *Lhx2* loss of function (*9*), as were most KLF-related footprints (Fig. 4G,
263 Auxiliary table 8,9).
264 NF-I, Sox2, KLF, Ascl1 and Hes5 showed a significant reduction in the mean
265 footprint scores at the center of their respective motifs (Fig. 5A-F). For example, 84.4%
266 of the NF-I footprints were lost in the E14 *Lhx2* mutant retinas, as were 93% of the Sox2
267 associated footprints and 87.6 % of the KLF-related footprints (Auxiliary table 11). The
268 mean footprint signal was also reduced in both E14 and P2 cKO samples (Auxiliary table
269 12). Among motifs that co-occur with Lhx2, a subset are recognized by TFs with known
270 or predicted pioneer function and show differential expression in early and late RPCs,
271 potentially priming closed chromatin for subsequent opening (Auxiliary table 14).

272 To better understand the interplay between chromatin organization and Lhx2-
273 dependent regulation of TF binding, we analyzed the nucleosome occupancy around the
274 TF binding sites. Nucleosome occupancy at the center of these motifs was altered in
275 Lhx2 cKO samples, suggesting that Lhx2 may facilitate or stabilize binding by these TFs
276 (Fig. 5G-L). Lhx2 competed for nucleosome occupancy in a manner like that previously
277 described for the non-canonical pioneer factor Nfib (*13*), with competition was more
278 pronounced at E14 than at P2 (Fig. 5G). In E14 and P2 *Lhx2* cKO retinas, an overall
279 increase in nucleosome coverage occurs, consistent with the reduction in chromatin
280 accessibility observed upon *Lhx2* loss of function. An increase in chromatin compaction
281 was also observed around the Sox2 motif center, although no difference in competition
282 could be detected in *Lhx2* cKO samples (Fig. 5H), suggesting that Sox2 is preferentially
283 recruited to nucleosome-free regions. NFI and KLF factors, as well as Hes5, competed
284 directly for nucleosome occupancy, although no difference was seen on motif centers for
285 NFI and KLF in *Lhx2* cKO samples (Fig. 5I-K). A reduction in chromatin compaction
286 associated with these NFI and KLF sites was seen in the E14 *Lhx2* cKO, while the
287 opposite effect was observed in the P2 *Lhx2* cKO, implying that NFI and KLF factors
288 mediate chromatin unfolding in late-stage RPCs. Finally, nucleosome occupancy by
289 Ascl1 was reduced in the *Lhx2* cKO retinas at both time points, with an effect on flanking
290 regions also observed at P2 (Fig. 5L).
291
292 **Discussion:**
293 In this study, we report a central role for Lhx2 in the global control of chromatin
294 accessibility in RPCs and identify possible mechanisms by which Lhx2 may regulate
295 gene expression over the course of retinal neurogenesis. First, an unbiased search for
296 oligonucleotides enriched in OCRs found in both early and late RPCs allowed the
297 identification of Lhx2 as the most overrepresented canonical transcription factor. The
298 Lhx2 *cis*-regulatory repertoire changes over the course of neurogenesis, where the
299 selective occupancy of retinal targets is reflected into gene variations upon Lhx2 loss of
300 function, with preferential targeting of genes specifically expressed in early retinal
301 progenitors at embryonic stages, late progenitors and postnatal stages and a lower
302 representation of Muller glia and anterodorsal hypothalamus related genes sets. Changes
303 in gene expression and chromatin accessibility seen following loss of *Lhx2* function are
304 more modest in P2 RPCs than in E14 RPCs. While this fact may partially reflect more
305 extensive perdurance of Lhx2 expression and footprinting that is seen in at P2 RPCs (Fig.
306 4G), it may indicate that a more diverse, and partially *Lhx2*-independent, transcriptional
307 regulatory network regulates the activity and accessibility of these sequences in late-stage
308 RPCs.
309 Second, Lhx2 affects local and global chromatin accessibility in RPCs by
310 competition for nucleosome occupancy and through nucleosome remodeling. Lhx2
311 target sites are often coupled with active enhancer marks, or flanked by active promoter
312 regions within modules of accessible chromatin. *Lhx2* loss of function affects expression
313 of genes involved in the cell cycle checkpoints, DNA replication, axonal dystrophy and
314 the Notch signaling pathway, consistent with previous findings (*7, 9*).
315 Third, Lhx2-dependent regulation of both expression and function of pioneer TFs
316 in RPCs appears to contribute substantially to these changes in chromatin accessibility.
317 The regulation of known pioneer factors by Lhx2 is achieved both transcriptionally, as is

318 the case of Lhx2-dependent regulation of OCRs associated with putative cis-regulatory
319 elements associated with the *Sox2* locus, and by stabilizing the binding these TFs to their
320 target sites, as in the case of Sox9, Ascl1 and Hes5. A number of Lhx2-regulated
321 predicted pioneer factors show substantially elevated expression in late-stage RPCs.
322 These include *Hes5*, *Sox8/9*, and *Nfia*, which play important roles in the control of retinal
323 gliogenesis, a process that is dependent on *Lhx2* (*9*). *Ascl1*, which is also selectively
324 expressed in late-stage RPCs, plays an essential role in conferring neurogenic
325 competence on late-stage RPCs (*15*). In contrast, other direct Lhx2 targets such as *Pax6*
326 and *Sox2*, like Lhx2, both show broad temporal expression and perform different
327 functions in early and late-stage RPCs (*19-22*). Some of the known and candidate
328 pioneer factors that are selectively expressed in late-stage RPCs – such as *Nfia/b/x*,
329 *Sox8/9*, and *Klf4/9/13* -- may preserve chromatin in a highly compacted state at
330 embryonic states, while mediating chromatin unfolding later in development.
331 The precise molecular mechanism by which Lhx2 mediates these
332 developmentally dynamic changes in chromatin accessibility is unclear. However, recent
333 studies of Lhx2 function in developing cortex have identified multiple histone modifying
334 enzymes that interact with Lhx2 (*23*). Interestingly, these include NuRD complex
335 components that are associated with transcriptional repression and reduced chromatin
336 accessibility, such as LSD1, HDAC2, and RBBP4. In RPCs, Lhx2 may instead recruit
337 different sets of histone remodeling enzymes that promote chromatin accessibility, which
338 should be readily identifiable through biochemical analysis.
339
340
341

342 **Figure legends:**
343
344 **Fig 1. Pairwise comparison of ATAC-Seq data identifies OCRs from early and late**
345 **RPCs, with a broad overrepresentation of Lhx2-related motifs. (A)** Workflow
346 adopted for epigenomic profiling of RPC **(B)** Venn diagram represents open chromatin
347 regions identified in early and late-stage RPCs **(C)** Clustering of TF motifs based on
348 similarity (left, Jaspar 2016 non-redundant vertebrates core). The heatmap indicates
349 similarity for each pairwise comparison. The color bars along the axis represent different
350 TF subfamilies. Inset represents a homeobox cluster that was enriched in OCRs in RPCs
351 (scale= 1 node/pixel) **(D)** Relative expression by RNA-Seq for homeobox TFs **(E)**
352 Representative Lhx2-related logos are shown (MA0700.1, k-mer sig=300, e-value=1e-
353 300) **(F)** Binomial scoring of known motifs in chromatin regions accessible in early and
354 late RPCs **(G)** Lhx2 footprints from OCRs identified in early and late-stage RPCs **(H)**
355 Custom tracks of RPC-derived ATAC-Seq profiles and aged-matched Lhx2 ChIP-Seq
356 feature the *Vsx2* locus.
357
358 **Figure 2. Lhx2 regulates cell cycle genes and the Notch signaling pathway by**
359 **targeting promoters and non-coding elements in nucleosome-free regions associated**
360 **with active enhancers (A)** Lhx2 known motif density for E14 and P2 ChIP-Seq peaks.
361 (**B**) Lhx2 peak association with specific genomic regions ($\log_2$-fold enrichment). (**C-F**)
362 Enrichment for cell-specific gene expression patterns is shown for Lhx2 peaks that are
363 shared between E14 and P2 (C), associated with stage-specific Lhx2 peaks (D), and all
364 Lhx2 peaks detected at E14 (E) and P2 (F). RNA-Seq from age-matched Lhx2 cKO
365 retinas identifies Lhx2-dependent genes sets associated with at least one ChIP-Seq peak.
366 Asterisks refer to p-values after Bonferroni-Hochberg correction. **(G,H)** Heatmaps of raw
367 reads from Lhx2 ChIP-Seq peaks are plotted across nucleosome centered regions
368 identified using age-matched ATAC-Seq samples. Each row represents a 3 kb window
369 (1.5 kb each direction) centered at the maximum read pile-up. For Lhx2 motif occurrence
370 in OCRs, refer to Auxiliary table S8. Lhx2 motifs at open chromatin regions co-
371 directionally distribute with RNA-Seq raw reads. Metaprofiles of the class II enhancer-
372 associated H3K27ac marks were compiled at actively transcribed regions from the E14
373 GFP-positive (**G**) and P2 GFP-positive RPCs (**H**) (replicates in grey scale, background
374 in light grey). (**I,J**) Binomial scoring of Gene Ontology (GO) functional categories that
375 are enriched in genes associated with OCRs that overlap with Lhx2 peaks at E14 (**I**) or P2
376 (**J**).
377
378 **Figure 3**. **Binding sites for transcription factors with predicted pioneer function co-**
379 **occur with Lhx2 peaks (A,B)** Hierarchical clustering of Lhx2 ChIP-Seq regulatory
380 motifs and assigned representative logos are represented at E14 (**A**) and P2 (**B**)
381 (linkage=average; similarity threshold cor =0.6, ncor=0.4, w=5). The most enriched
382 cluster comprises Lhx2 and multiple close variations of the Lhx2 motif. **(C,D)** Known TF
383 motifs found preferentially enriched in either E14 or P2 Lhx2 ChIP-Seq peaks, and
384 identified by pairwise comparison, are shown. The percentages of Lhx2 ChIP-Seq peaks
385 that contain the indicated motif are listed, along with the background frequency for the
386 motif in question. The relative expression level of the corresponding TF mRNA at E14 or
387 P2 is indicated by the red/blue color gradient.

388
389 **Figure 4**. **Lhx2 affects local and global chromatin accessibility in RPCs by**
390 **regulating expression and DNA binding of pioneer TFs (A,B)** ATAC-Seq read
391 distribution plotted relative to the center of Lhx2 footprints, and average inflection depth
392 for these footprints in both controls and *Lhx2* mutants. **(C,D)** Heat maps of raw reads
393 from Lhx2 ChIP-Seq and age-matched OCRs, identified by ATAC-Seq of purified RPCs,
394 are plotted in a 3kb window relative to the center of OCRs from control and *Lhx2* cKO
395 retinas, with the difference in signal between these two conditions shown at right. **(E,F)**
396 Paired variation of RNA-Seq signal and local chromatin accessibility in E14 (**E**) and P2
397 (**F**) control and *Lhx2* cKO samples are shown for loci encoding TFs that are regulated by
398 Lhx2. Individual representative TFs are highlighted by arrows. **(G)** Footprint counts for
399 individual TFs with predicted pioneer activity that are associated with Lhx2 target sites
400 are shown in both control and *Lhx2* mutants from E14 and P2 retina.

401 **Fig.5. Metaprofiles of footprinting and nucleosome occupancy data reveal**
402 **developmentally dynamic regulation by Lhx2 of DNA binding by RPC-expressed**
403 **TFs with predicted pioneer activity. (A-F)** ATAC-Seq average cut profiles feature
404 footprints for Lhx2, Sox2, NF-I (Nfia/b/x), Klf4/9, Hes5 and Ascl1 from both E14 and P2
405 control and *Lhx2* cKO samples. (**G-L**) Nucleosome occupancy at the center of each motif
406 is reported for Lhx2, Sox2, NF-I (Nfia/b/x), Klf4/9, Hes5, and Ascl1 in both E14 and P2
407 control and *Lhx2* cKO samples.

408
409
410

411 **Materials and Methods:**
412
413 **Experimental design:**
414 The experimental pipeline involves the generation and integration of high-
415 throughput sequencing libraries from purified fractions of murine retinal progenitors cells
416 (RPCs). Fluorescence-activated cell sorting (FACS) was adopted to isolate RPCs from
417 mice expressing the RPC-specific *Chx10-Cre:eGFP* transgene (*11*) from embryonic day
418 (E)14 and postnatal day (P)2 retina, representing early and late stages of RPC
419 competence, respectively. Flow-sorted cells were profiled by RNA-Seq and ATAC-Seq,
420 relying on direct *in vitro* transposition of sequencing adapters into native chromatin.
421 ChIP-Seq was then performed on a select transcription factor candidate, identified in the
422 preliminary screening of the ATAC-Seq derived open chromatin regions. ChIP-Seq
423 binding sites for the candidate factor were ultimately integrated and compared with the
424 age-matched RNA-Seq and ATAC-Seq profiles from control and *Lhx2* cKO retinas. For
425 each experimental condition involving a point-source factor and broad regions (ChIP-Seq
426 and ATAC-Seq, respectively) a minimum of 20 Millions uniquely mappable reads or ≥10
427 millions uniquely mappable reads for each biological replicate were collected, according
428 to the ENCODE's guidelines. For point-source datasets, non-redundant mapped reads
429 were retained for downstream analysis.
430
431 **Animals**
432 CD1 mice of either sex were euthanized at embryonic day 14 (E14) and postnatal
433 day 2 (P2) according to Johns Hopkins IACUC-approved protocols. Timed pregnant CD-
434 1 mice were obtained from Charles River Laboratories. *Chx10-Cre:GFP* mice (*11*) were
435 purchased from the Jackson Laboratories. Retinas were freshly dissected, incubated in a
436 suspension of papain and DNAse for 30 min at 37 C, inactivated with bovine serum
437 albumin, resuspended in equilibrated Earle's balanced salt solution and subject to
438 fluorescence activated cell sorting (FACS) to 98-99% purity, with viability assessed by
439 propidium iodide exclusion. Cell fractions were collected on poly-D-lysine coated slides,
440 fixed in 4% paraformaldehyde for 10 min, permeabilized in TritonX-100 and stained for
441 Chx10 (Cat.# X1179P, Exalpha), GFP (Cat.# 600-101-215, Rockland), Ki67 (Cat.# RM-
442 9106-S1,Thermo Scientific) or Ccnd1 (Cat.# SC-450, Santa Cruz). The brightest fraction,
443 which consistently showed 4-fold higher mean intensity for GFP relative to the dim
444 fraction, was always retained for subsequent processing and hereafter referred to as GFP-
445 positive, RPC-enriched fraction. RNA-Seq analysis revealed a substantial enrichment of
446 RPC-specific markers in the GFP-positive fraction relative to the age-matched GFP-
447 negative fraction. Examples of this include (number indicates level in GFP-positive
448 relative to age-matched GFP-negative cells): Ccnd1, E14=22.9, P2=2.9; Mki67,
449 E14=19.7, P2=3.1; Vsx2, E14=47.7, P2=2.8; Fgf15, E14=45.4, P2=3.4.
450 *Lhx2* conditional embryonic knockouts were obtained by crossing *Chx10-Cre:GFP*
451 with $Lhx2^{lox/lox}$ mice, and harvesting at E14 (*24*). Postnatal *Lhx2* knockouts were
452 generated by electroporation of pCAG-Cre-GFP construct into P0.5 wild type CD1
453 animals or $Lhx2^{lox/lox}$ animals. Retinas were harvested at P2, dissociated, and GFP-
454 positive electroporated cells were isolated by FACS. Overall electroporation efficiency
455 was 2-3%.
456

457 **ATAC-Seq, RNA-Seq, and ChIP-Seq analysis**
458 Chromatin derived from flow-sorted *Chx10-Cre*GFP+ve and GFP-ve retinal
459 fractions was processed as previously described (*25*). Briefly, chromatin was extracted
460 and processed for Tn5 mediated tagmentation and adapter incorporation, according to the
461 Manufacturer's protocol (Nextera DNA sample preparation kit, Illumina®) at 37 C for
462 30 min. Reduced-cycle amplification was carried out in presence of compatibly indexed
463 sequencing adapters. The quality of the libraries was assessed by fluorometric DNA
464 incorporation based assay (Thermo Fisher Scientific™) and automated capillary
465 electrophoresis (Agilent Technologies, Inc.) and up to 4 samples per lane were pooled
466 and run as 50 bp paired ends on a HiSeq2500 Illumina sequencer.
467 RNA was processed with Qiagen RNAeasy Mini kit, subject to DNAse digestion,
468 and samples with a minimum RNA integrity number (RIN) of 7 were further processed
469 for sequencing. Libraries were prepared using Illumina TruSeq RNA Sample kit
470 (Illumina, San Diego, CA) following manufacturer's recommended procedure. Briefly,
471 total RNA was denatured at 65°C for 5 minutes, cooled on ice, purified and incubated at
472 80°C for two minutes. The eluted mRNA was fragmented at 94°C for 8 min and
473 converted to double stranded cDNA, end repaired, A-tailed, and ligated with indexed
474 adaptors and run on a MiSeq Illumina sequencer. The quality of the libraries was
475 assessed by fluorometric RNA incorporation based assay (Thermo Fisher Scientific™)
476 and automated capillary electrophoresis (Agilent Technologies, Inc).
477 ChIP was performed as described previously (*26*). Whole dissected retinas were
478 dissociated in a collagenase I suspension, cross-linked in 1% formaldehyde for 15 min,
479 and quenched in 125 mM glycine. The extracted nuclei were sheared to produce 100–500
480 bp fragments by means of probe sonication. Chromatin was immunoprecipitated with
481 anti-Lhx2 (Cat.# SC-19344, Santa Cruz Biotechnology), rabbit anti-H3K27Ac ((Cat.#
482 ab4729, Abcam), or the corresponding isotype controls (Abcam); retained on agarose
483 beads (Invitrogen), washed and purified by organic extraction. Success of both anti-Lhx2
484 and anti-H3K27Ac ChIP was confirmed using SYBR qRT-PCR (Agilent Technologies)
485 that performed using previously described primer pairs that amplify both cis-regulatory
486 regions that exhibit Lhx2-binding sites and syntenic unbound regions (*9*). Libraries were
487 processed according to the manufacturer's protocol (TruSeq Nano DNA Library Prep
488 Kit). The quality of the libraries was assessed by fluorometric DNA incorporation based
489 assay (Thermo-Fisher) and automated capillary electrophoresis (Agilent Technologies)
490 and libraries (100-150bp single read, paired ends) were run on a HiSeq2500 Illumina
491 sequencer.
492
493 **Peak calling and retinal gene ontology analysis**
494 RNA-Seq reads were aligned to the mouse transcriptome (mm9 UCSC build) using
495 Tophat2 (*27*), and differentially expressed genes (DEGs) were identified by Cuffdiffs
496 (*28*), imposing a cutoff q-val = 0.05 for pairwise comparison.
497 Bowtie2 was used for ChIP-Seq and ATAC-Seq reads alignment on the mouse
498 genome (mm9) (*29*). Uniquely mappable reads from ChIP-Seq were retained for peak
499 calling by MACSs (*30*) (band width = 300, mfold =5, 50, d = 200, max tags per position
500 =1, min FDR q-val cutoff = 1E-02, λ = 1000-10000 bp).
501 Open chromatin regions were identified in ATAC-Seq data using MACS2 (*30*).
502 Correlations between open chromatin states was identified by pairwise comparison of

503 normalized reads counts in E14 and P2 GFP-positive flow-sorted retinal progenitor cells
504 was done with Jaccard (*31*). The Jaccard index was estimated as the number of peaks
505 that overlap between two peak files, divided by the union of the two files. Footprints and
506 nucleosomes were identified as described previously (*12, 32, 33*). Annotation was done
507 by proximity to the closest transcription start site.
508 High-confidence ChIP-Seq peaks were identified from at least 2 experimental
509 replicates (Poisson p-val threshold = 0.0001, min FE=4, FDR=0.001, max tags per
510 position =1, normalization to input or isotype control) and subject to comparison with
511 ATAC-Seq peaks (hypergeometric test, ln p-val) where co-occurrence was defined by
512 physical overlap, allowing a max distance of 20 bp from peak summits over 3000 bp for
513 confocality. For direct comparison of ChIP-Seq and ATAC-Seq data, high confidence
514 peaks with the highest differential in accessibility between nucleosomal units and
515 flanking nucleosomes-free, transposon-accessible regions, and a minimum distance of
516 300bp were identified from two ATAC-Seq replicates.
517 Genes that showed enriched or specific expression in retinal progenitors were
518 identified generated based on the RNA-Seq data generated in this study (for RPC-specific
519 expression). Genes that showed cell-type enriched or specific expression in adult retina
520 were extracted from other published studies using either RNA-Seq or microarray-based
521 analysis of isolated cells, or *in situ* hybridization or immunohistochemistry of intact
522 retina. Association of ChIP-Seq peaks with genes expressed in specific retinal cell types
523 was evaluated by Fisher's exact test and corrected for multiple hypothesis testing
524 (Bonferroni).
525
526 **Motif enrichment analysis**
527 Hierarchical clustering of probabilistically assigned motifs (Jaspar 2016 non
528 redundant core vertebrates) (*34*) was done with the following parameters
529 (linkage=average; similarity threshold cor =0.6, ncor=0.4, rl=5, where ncor is Pearson's
530 correlation (cor) by relative alignment length (rl) divided by the overall alignment length.
531 (e-val= pval * enriched oligos) (*35*).
532 For ChIP-Seq analysis, binomial probability analysis of TF motifs was calculated in
533 the overall set of high confidence peaks (minimum of 2 normalized replicates). The
534 regions size was empirically determined and motifs were found by cumulative binomial
535 distribution of known position weight matrices assuming a random representation of
536 decamers. Motif finding was performed on the repeat-masked mm9 murine genome,
537 optimized for the top enriched 20 putative motifs, randomized and repeated twice to
538 estimate FDR (*36*).
539
540 **Footprint analysis**
541 FIMO (*37*) was used to scan the genome for candidate binding sites of different
542 transcription factors. We then used BPAC to identify the actual binding sites among these
543 candidate sites (*33*). The number of active binding sites was analyzed at both E14 and
544 P2. Genome-wide changes in footprint counts and nucleosome occupancy for individual
545 transcription factor motifs were estimated for all candidate TF binding sites at both E14
546 and P2. Nucleosome occupancy was estimated using NucleoATAC (*32*). Footprints
547 scores were calculated for Lhx2, Klf9/13, NF-I (Nfia/b/x), and Sox2. T-test on mean

548 footprints scores distributed 200bp within of motif site centers were calculated for the
549 paired control and *Lhx2* cKO conditions.
550
551 **Statistical analysis**
552 Co-occurrence statistics for point-source and broad regions of interest was
553 computed by hypergeometric test with a default minimum overlap of 1 bp, unless
554 otherwise specified. Coverage was adopted as reproducibility metrics for ChIP-Seq and
555 ATAC-Seq experimental replicates (fraction reads/10^7 non-redundant uniquely mappable
556 reads) and FPKM for RNA-Seq, where correlation was reported by Pearson's or
557 Spearman's coefficient. Binomial probability analysis of regulatory transcription factors
558 motifs was applied genome wide to identify enriched position weight matrices (PWMs)
559 and clustered by average linkage. Fisher's exact test was adopted to compute gene
560 ontology enrichment and corrected for multiple hypothesis (Bonferroni-Hochberg) and
561 RNA-Seq derived gene sets from flow sorted retinal cell fractions and control versus
562 experimental conditions (*Lhx2* cKO) with q-val (FDR) < 0.05 were retained for
563 downstream analysis, unless otherwise specified. Two-tailed t-test was adopted to
564 compare the average footprint counts for candidate pioneer factors in control and *Lhx2*
565 cKO samples.
566
567
568 **Acknowledgments**
569

570 **General**: We thank Hongkai Ji for help with statistical analysis, Anand Venkataraman
571 and Wendy Yap for comments on the manuscript, Hong Wang for help with dissections,
572 the JHMI Deep Sequencing and Microarray core, the JHU GRCF High Throughput
573 Sequencing facility, and the JHSPH Flow Cytometry and Cell Sorting Core Facility.

574 **Funding:** This work was supported by the National Institutes of Health (R01EY020560
575 to S.B., R01EY024580 and R01EY023188 to J.Q.)

576 **Author contributions**: C.Z. carried out experiments and data analysis. S.L, J.W. and J.Q
577 carried out analysis. C.Z., J.Q. and S.B. conceived the study and wrote the paper.

578 **Competing interests:** The authors declare no conflict of interest.

579 **Data and materials availability:** Auxiliary data and files have been deposited in GEO,
580 Accession Number GSE99818. All data needed to evaluate the conclusions in the paper
581 are present in the paper and/or the Supplementary Materials. Additional data is available
582 from the authors upon request.
583
584 **References:**
585 1. F. J. Livesey, C. L. Cepko, Vertebrate neural cell-fate determination: lessons from
586 the retina. *Nat Rev Neurosci* **2**, 109-118 (2001).
587 2. M. Kohwi, C. Q. Doe, Temporal fate specification and neural progenitor
588 competence during development. *Nat Rev Neurosci* **14**, 823-838 (2013).
589 3. J. Elliott, C. Jolicoeur, V. Ramamurthy, M. Cayouette, Ikaros confers early
590 temporal competence to mouse retinal progenitor cells. *Neuron* **60**, 26-39 (2008).

591 4. P. Mattar, J. Ericson, S. Blackshaw, M. Cayouette, A conserved regulatory logic
592 controls temporal identity in mouse neural progenitors. *Neuron* **85**, 497-504
593 (2015).
594 5. Z. Yang, K. Ding, L. Pan, M. Deng, L. Gan, Math5 determines the competence
595 state of retinal ganglion cell progenitors. *Dev Biol* **264**, 240-254 (2003).
596 6. S. Li *et al.*, Foxn4 controls the genesis of amacrine and horizontal cells by retinal
597 progenitors. *Neuron* **43**, 795-807 (2004).
598 7. P. J. Gordon *et al.*, Lhx2 balances progenitor maintenance with neurogenic output
599 and promotes competence state progression in the developing retina. *J Neurosci*
600 **33**, 12197-12207 (2013).
601 8. J. de Melo, B. S. Clark, S. Blackshaw, Multiple intrinsic factors act in concert
602 with Lhx2 to direct retinal gliogenesis. *Scientific reports* **6**, 32757 (2016).
603 9. J. de Melo *et al.*, Lhx2 Is an Essential Factor for Retinal Gliogenesis and Notch
604 Signaling. *J Neurosci* **36**, 2391-2405 (2016).
605 10. A. Roy *et al.*, LHX2 is necessary for the maintenance of optic identity and for the
606 progression of optic morphogenesis. *J Neurosci* **33**, 6877-6884 (2013).
607 11. S. Rowan, C. L. Cepko, Genetic analysis of the homeodomain transcription factor
608 Chx10 in the retina using a novel multifunctional BAC transgenic mouse reporter.
609 *Dev Biol* **271**, 388-402 (2004).
610 12. R. I. Sherwood *et al.*, Discovery of directional and nondirectional pioneer
611 transcription factors by modeling DNase profile magnitude and shape. *Nat*
612 *Biotechnol* **32**, 171-178 (2014).
613 13. S. K. Denny *et al.*, Nfib Promotes Metastasis through a Widespread Increase in
614 Chromatin Accessibility. *Cell* **166**, 328-342 (2016).
615 14. X. Liu *et al.*, Yamanaka factors critically regulate the developmental signaling
616 network in mouse embryonic stem cells. *Cell Res* **18**, 1177-1189 (2008).
617 15. J. A. t. Brzezinski, E. J. Kim, J. E. Johnson, T. A. Reh, Ascl1 expression defines a
618 subpopulation of lineage-restricted progenitors in the mammalian retina.
619 *Development* **138**, 3519-3531 (2011).
620 16. M. Hojo *et al.*, Glial cell fate specification modulated by the bHLH gene Hes5 in
621 mouse retina. *Development* **127**, 2515-2522 (2000).
622 17. A. Muto, A. Iida, S. Satoh, S. Watanabe, The group E Sox genes Sox8 and Sox9
623 are regulated by Notch signaling and are required for Muller glial cell
624 development in mouse retina. *Exp Eye Res* **89**, 549-558 (2009).
625 18. M. A. Lodato *et al.*, SOX2 co-occupies distal enhancer elements with distinct
626 POU factors in ESCs and NPCs to specify cell state. *PLoS Genet* **9**, e1003288
627 (2013).
628 19. T. Marquardt *et al.*, Pax6 is required for the multipotent state of retinal progenitor
629 cells. *Cell* **105**, 43-55 (2001).
630 20. Y. P. Lin, Y. Ouchi, S. Satoh, S. Watanabe, Sox2 plays a role in the induction of
631 amacrine and Muller glial cells in mouse retinal progenitor cells. *Invest*
632 *Ophthalmol Vis Sci* **50**, 68-74 (2009).
633 21. D. Matsushima, W. Heavner, L. H. Pevny, Combinatorial regulation of optic cup
634 progenitor cell fate by SOX2 and PAX6. *Development* **138**, 443-454 (2011).

635 22. L. A. Remez *et al.*, Pax6 is essential for the generation of late-born retinal neurons
636 and for inhibition of photoreceptor-fate during late stages of retinogenesis. *Dev*
637 *Biol* **432**, 140-150 (2017).
638 23. B. Muralidharan *et al.*, LHX2 Interacts with the NuRD Complex and Regulates
639 Cortical Neuron Subtype Determinants Fezf2 and Sox11. *J Neurosci* **37**, 194-203
640 (2017).
641 24. V. S. Mangale *et al.*, Lhx2 selector activity specifies cortical identity and
642 suppresses hippocampal organizer fate. *Science* **319**, 304-309 (2008).
643 25. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf,
644 Transposition of native chromatin for fast and sensitive epigenomic profiling of
645 open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*
646 **10**, 1213-1218 (2013).
647 26. Y. Shang, X. Hu, J. DiRenzo, M. A. Lazar, M. Brown, Cofactor dynamics and
648 sufficiency in estrogen receptor-regulated transcription. *Cell* **103**, 843-852 (2000).
649 27. D. Kim *et al.*, TopHat2: accurate alignment of transcriptomes in the presence of
650 insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).
651 28. C. Trapnell *et al.*, Differential gene and transcript expression analysis of RNA-seq
652 experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-578 (2012).
653 29. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat*
654 *Methods* **9**, 357-359 (2012).
655 30. Y. Zhang *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**,
656 R137 (2008).
657 31. A. Rau, M. Gallopin, G. Celeux, F. Jaffrezic, Data-based filtering for replicated
658 high-throughput transcriptome sequencing experiments. *Bioinformatics* **29**, 2146-
659 2152 (2013).
660 32. A. N. Schep *et al.*, Structured nucleosome fingerprints enable high-resolution
661 mapping of chromatin architecture within regulatory regions. *Genome Res* **25**,
662 1757-1770 (2015).
663 33. S. Liu *et al.*, BPAC: A universal model for prediction of tran-scription factor
664 binding sites based on chromatin accessibility. *BMC Bioinformatics* **18**, 355
665 (2017).
666 34. A. Mathelier *et al.*, JASPAR 2016: a major expansion and update of the open-
667 access database of transcription factor binding profiles. *Nucleic Acids Res* **44**,
668 D110-115 (2016).
669 35. J. van Helden, Regulatory sequence analysis tools. *Nucleic Acids Res* **31**, 3593-
670 3596 (2003).
671 36. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors
672 prime cis-regulatory elements required for macrophage and B cell identities. *Mol*
673 *Cell* **38**, 576-589 (2010).
674 37. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: scanning for occurrences of a
675 given motif. *Bioinformatics* **27**, 1017-1018 (2011).
676

Fig.1

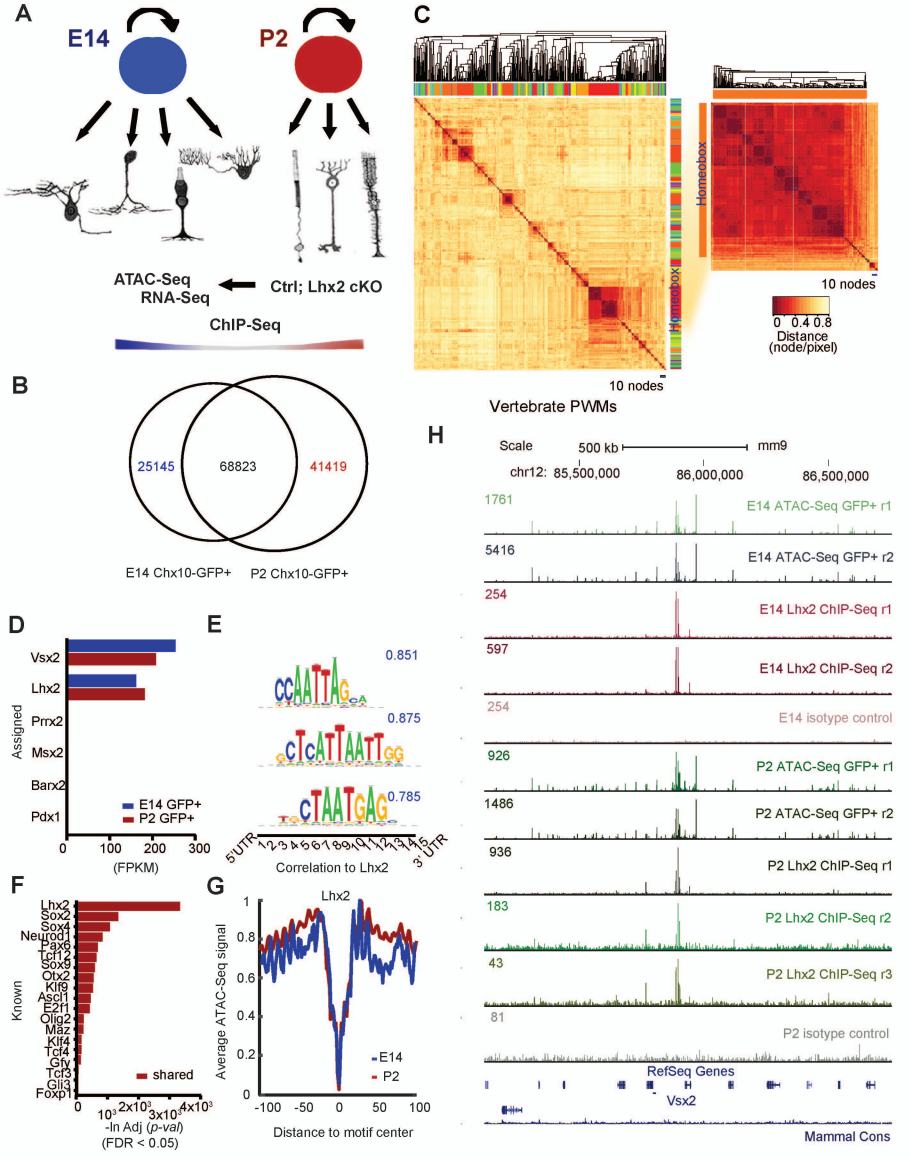

A
E14
P2
ATAC-Seq
RNA-Seq
Ctrl; Lhx2 cKO
ChIP-Seq
B
25145
68823
41419
E14 Chx10-GFP+
P2 Chx10-GFP+
C
Homeobox
10 nodes
0 0.4 0.8
Distance
(node/pixel)
Vertebrate PWMs
D
Vsx2
Lhx2
Prrx2
Msx2
Barx2
Pdx1
Assigned
E14 GFP+
P2 GFP+
0
100
200
300
(FPKM)
E
0.851
0.875
0.785
5'UTR
3'UTR
Correlation to Lhx2
F
Lhx2
Sox2
Sox4
Neurod1
Pax6
Tcf12
Sox3
Otx2
Klf9
Ascl1
E2f1
Olig2
Maz
Klf4
Tcf4
Gfy
Tcf3
Gli3
Foxp1
Known
shared
-ln Adj (p-val)
(FDR < 0.05)
G
Lhx2
Average ATAC-Seq signal
E14
P2
Distance to motif center
H
Scale
500 kb
mm9
chr12: 85,500,000
86,000,000
86,500,000
1761
E14 ATAC-Seq GFP+ r1
5416
E14 ATAC-Seq GFP+ r2
254
E14 Lhx2 ChIP-Seq r1
597
E14 Lhx2 ChIP-Seq r2
254
E14 isotype control
926
P2 ATAC-Seq GFP+ r1
1486
P2 ATAC-Seq GFP+ r2
936
P2 Lhx2 ChIP-Seq r1
183
P2 Lhx2 ChIP-Seq r2
43
P2 Lhx2 ChIP-Seq r3
81
P2 isotype control
RefSeq Genes
Vsx2
Mammal Cons

Fig.2

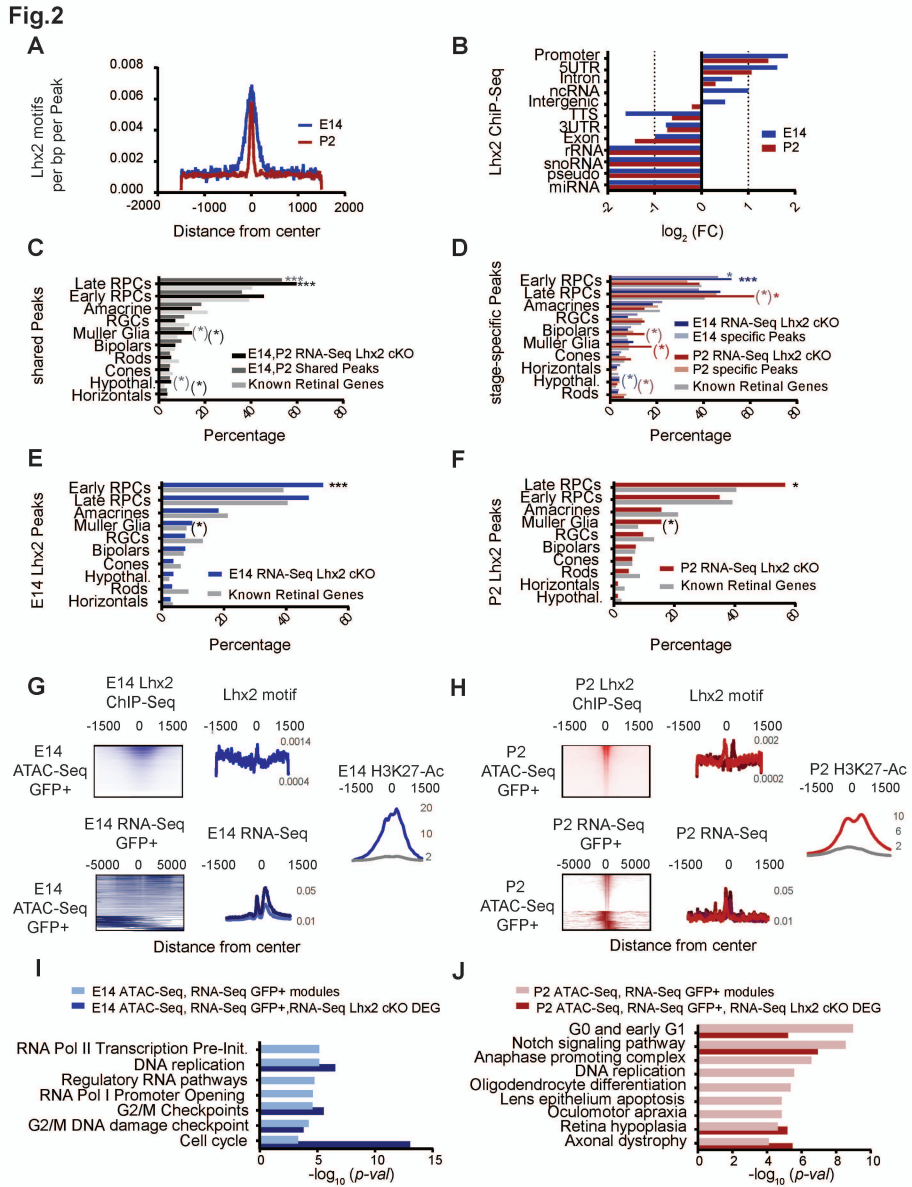

A
Lhx2 motifs per bp per Peak
0.008
0.006
0.004
0.002
0.000
-2000
-1000
0
1000
2000
Distance from center
E14
P2
B
Lhx2 ChIP-Seq
Promoter
5UTR
Intron
ncRNA
Intergenic
TTS
3UTR
Exon
rRNA
snoRNA
pseudo
miRNA
E14
P2
log2 (FC)
C
shared Peaks
Late RPCs
Early RPCs
Amacrine
RGCs
Muller Glia
Bipolars
Rods
Cones
Hypothal.
Horizontals
E14,P2 RNA-Seq Lhx2 cKO
E14,P2 Shared Peaks
Known Retinal Genes
Percentage
D
stage-specific Peaks
Early RPCs
Late RPCs
Amacrines
RGCs
Bipolars
Muller Glia
Cones
Horizontals
Hypothal.
Rods
E14 RNA-Seq Lhx2 cKO
E14 specific Peaks
P2 RNA-Seq Lhx2 cKO
P2 specific Peaks
Known Retinal Genes
Percentage
E
E14 Lhx2 Peaks
Early RPCs
Late RPCs
Amacrines
Muller Glia
RGCs
Bipolars
Cones
Hypothal.
Rods
Horizontals
E14 RNA-Seq Lhx2 cKO
Known Retinal Genes
Percentage
F
P2 Lhx2 Peaks
Late RPCs
Early RPCs
Amacrines
Muller Glia
RGCs
Bipolars
Cones
Rods
Horizontals
Hypothal.
P2 RNA-Seq Lhx2 cKO
Known Retinal Genes
Percentage
G
E14 Lhx2 ChIP-Seq
Lhx2 motif
E14 ATAC-Seq GFP+
E14 H3K27-Ac
E14 RNA-Seq GFP+
E14 RNA-Seq
E14 ATAC-Seq GFP+
Distance from center
H
P2 Lhx2 ChIP-Seq
Lhx2 motif
P2 ATAC-Seq GFP+
P2 H3K27-Ac
P2 RNA-Seq GFP+
P2 RNA-Seq
P2 ATAC-Seq GFP+
Distance from center
I
E14 ATAC-Seq, RNA-Seq GFP+ modules
E14 ATAC-Seq, RNA-Seq GFP+,RNA-Seq Lhx2 cKO DEG
RNA Pol II Transcription Pre-Init.
DNA replication
Regulatory RNA pathways
RNA Pol I Promoter Opening
G2/M Checkpoints
G2/M DNA damage checkpoint
Cell cycle
-log10 (p-val)
J
P2 ATAC-Seq, RNA-Seq GFP+ modules
P2 ATAC-Seq, RNA-Seq GFP+, RNA-Seq Lhx2 cKO DEG
G0 and early G1
Notch signaling pathway
Anaphase promoting complex
DNA replication
Oligodendrocyte differentiation
Lens epithelium apoptosis
Oculomotor apraxia
Retina hypoplasia
Axonal dystrophy
-log10 (p-val)

Fig.3

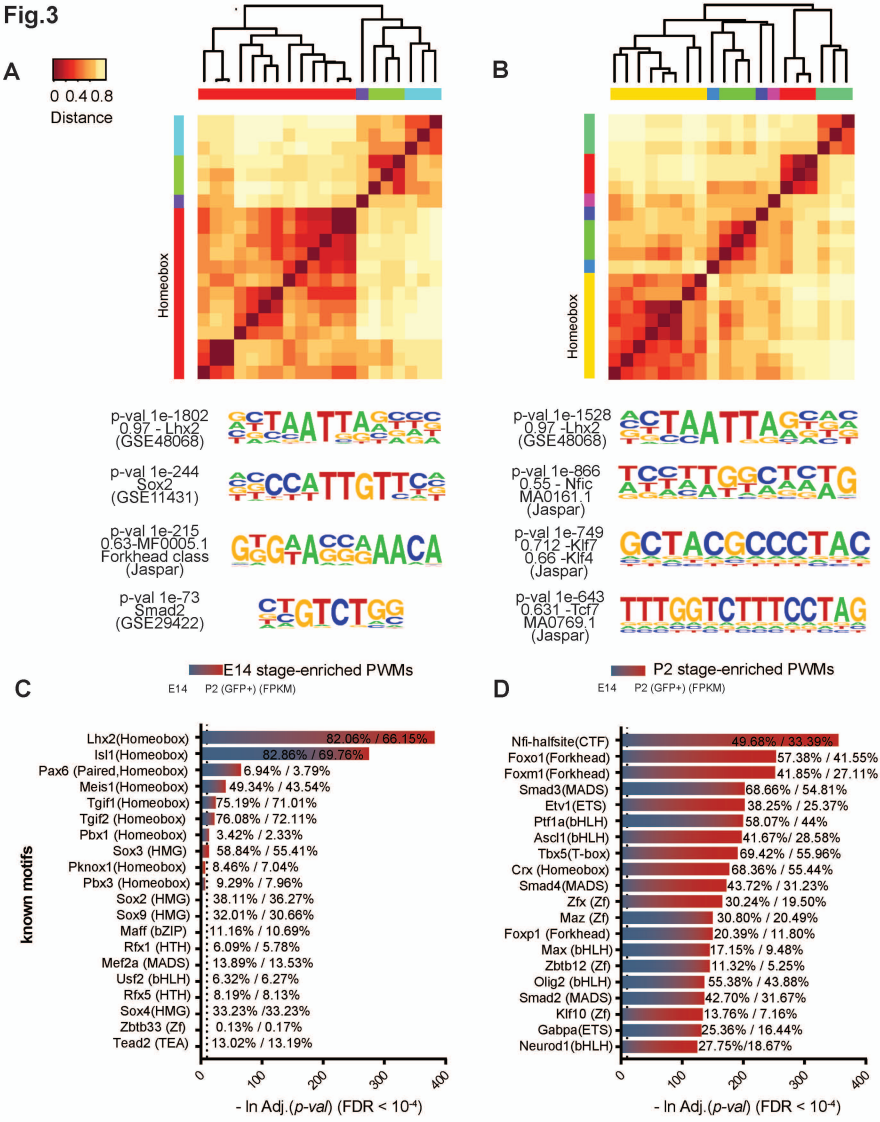

A
0 0.4 0.8
Distance
Homeobox
B
Homeobox
p-val 1e-1802
0.97 - Lhx2
(GSE48068)
p-val 1e-244
Sox2
(GSE11431)
p-val 1e-215
0.63-MF0005.1
Forkhead class
(Jaspar)
p-val 1e-73
Smad2
(GSE29422)
p-val 1e-1528
0.97 -Lhx2
(GSE48068)
p-val 1e-866
0.55 - Nfic
MA0161.1
(Jaspar)
p-val 1e-749
0.712 -Klf7
0.66 -Klf4
(Jaspar)
p-val 1e-643
0.631 -Tcf7
MA0769.1
(Jaspar)
C
E14 stage-enriched PWMs
E14 P2 (GFP+) (FPKM)
known motifs
Lhx2(Homeobox) 82.06% / 66.15%
Isl1(Homeobox) 82.86% / 69.76%
Pax6 (Paired,Homeobox) 6.94% / 3.79%
Meis1(Homeobox) 49.34% / 43.54%
Tgif1(Homeobox) 75.19% / 71.01%
Tgif2 (Homeobox) 76.08% / 72.11%
Pbx1 (Homeobox) 3.42% / 2.33%
Sox3 (HMG) 58.84% / 55.41%
Pknox1(Homeobox) 8.46% / 7.04%
Pbx3 (Homeobox) 9.29% / 7.96%
Sox2 (HMG) 38.11% / 36.27%
Sox9 (HMG) 32.01% / 30.66%
Maff (bZIP) 11.16% / 10.69%
Rfx1 (HTH) 6.09% / 5.78%
Mef2a (MADS) 13.89% / 13.53%
Usf2 (bHLH) 6.32% / 6.27%
Rfx5 (HTH) 8.19% / 8.13%
Sox4(HMG) 33.23% /33.23%
Zbtb33 (Zf) 0.13% / 0.17%
Tead2 (TEA) 13.02% / 13.19%
0 100 200 300 400
- ln Adj.(p-val) (FDR < 10-4)
D
P2 stage-enriched PWMs
E14 P2 (GFP+) (FPKM)
Nfi-halfsite(CTF) 49.68% / 33.39%
Foxo1(Forkhead) 57.38% / 41.55%
Foxm1(Forkhead) 41.85% / 27.11%
Smad3(MADS) 68.66% / 54.81%
Etv1(ETS) 38.25% / 25.37%
Ptf1a(bHLH) 58.07% / 44%
Ascl1(bHLH) 41.67%/ 28.58%
Tbx5(T-box) 69.42% / 55.96%
Crx (Homeobox) 68.36% / 55.44%
Smad4(MADS) 43.72% / 31.23%
Zfx (Zf) 30.24% / 19.50%
Maz (Zf) 30.80% / 20.49%
Foxp1 (Forkhead) 20.39% / 11.80%
Max (bHLH) 17.15% / 9.48%
Zbtb12 (Zf) 11.32% / 5.25%
Olig2 (bHLH) 55.38% / 43.88%
Smad2 (MADS) 42.70% / 31.67%
Klf10 (Zf) 13.76% / 7.16%
Gabpa(ETS) 25.36% / 16.44%
Neurod1(bHLH) 27.75%/18.67%
0 100 200 300 400
- ln Adj.(p-val) (FDR < 10-4)

Fig.4

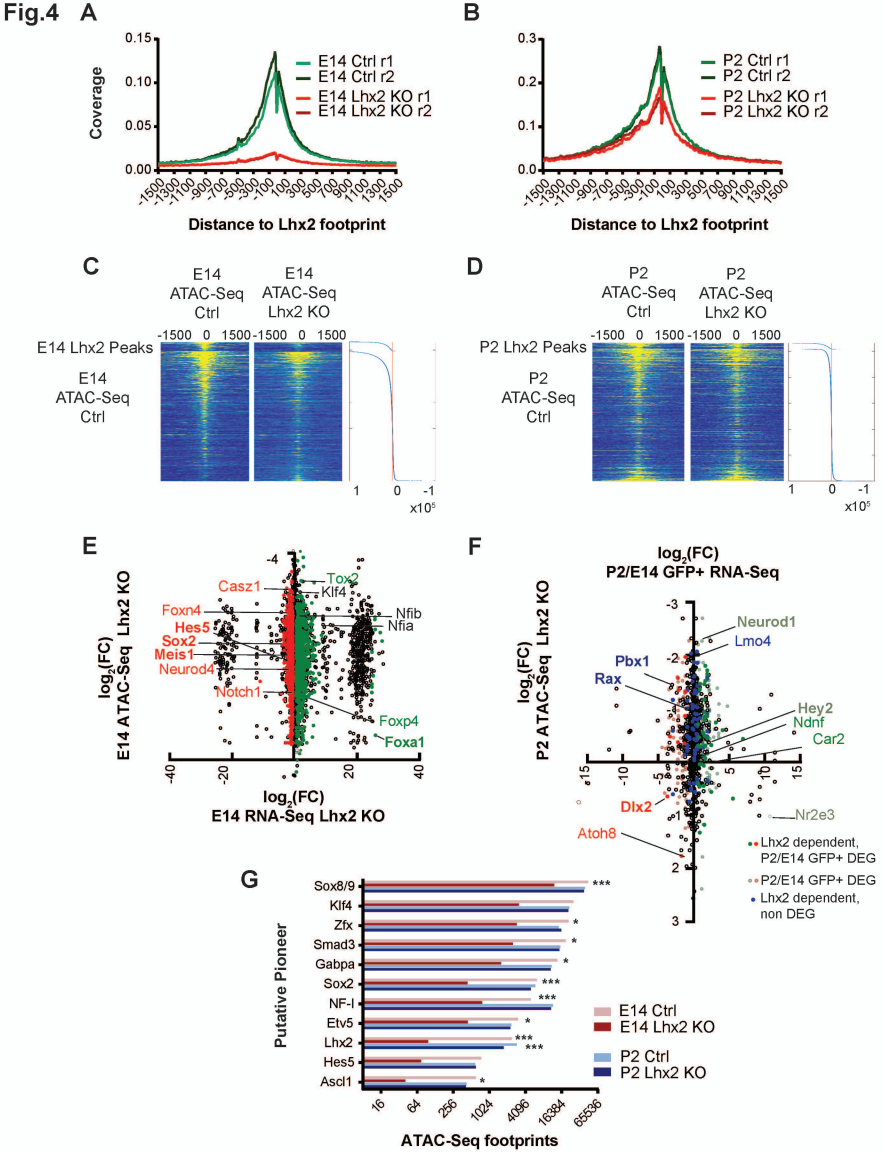
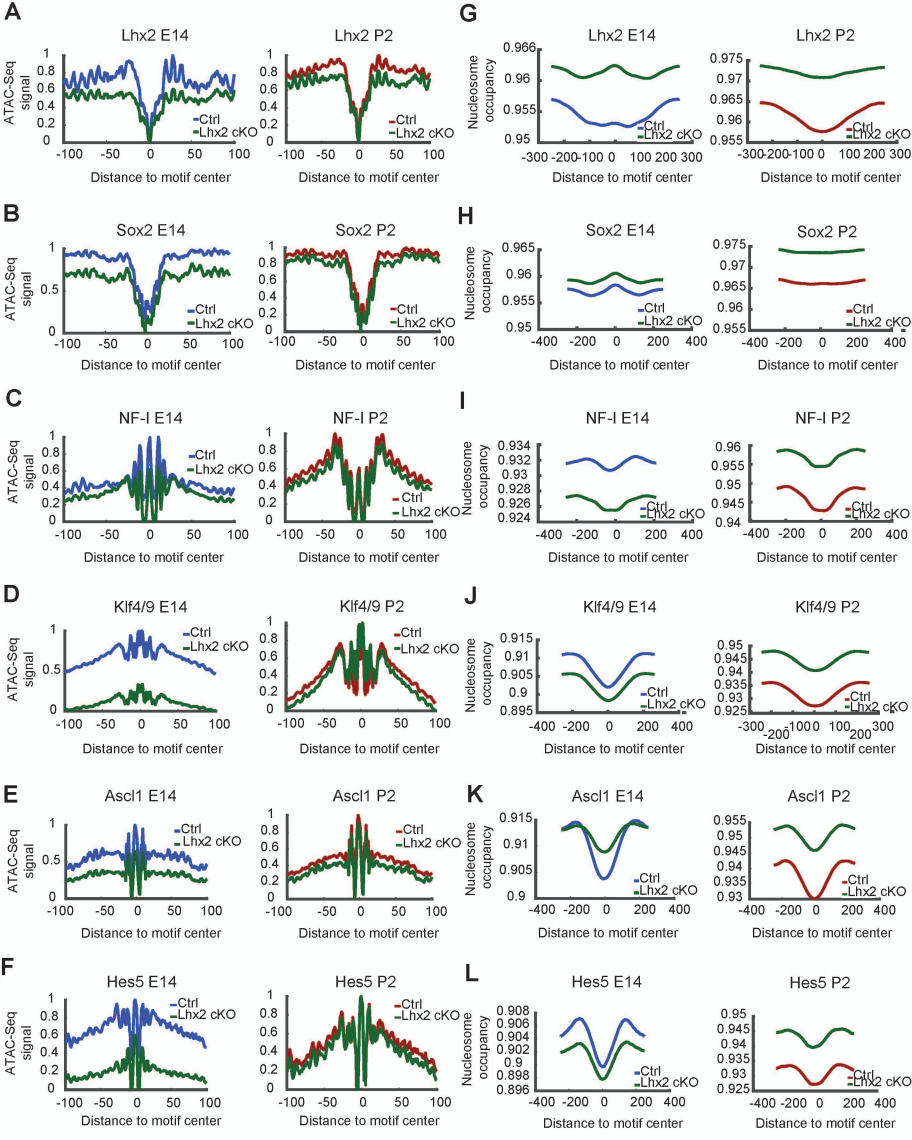

A
Coverage
E14 Ctrl r1
E14 Ctrl r2
E14 Lhx2 KO r1
E14 Lhx2 KO r2
Distance to Lhx2 footprint
B
P2 Ctrl r1
P2 Ctrl r2
P2 Lhx2 KO r1
P2 Lhx2 KO r2
Distance to Lhx2 footprint
C
E14 ATAC-Seq Ctrl
E14 ATAC-Seq Lhx2 KO
E14 Lhx2 Peaks
E14 ATAC-Seq Ctrl
D
P2 ATAC-Seq Ctrl
P2 ATAC-Seq Lhx2 KO
P2 Lhx2 Peaks
P2 ATAC-Seq Ctrl
E
log2(FC) E14 ATAC-Seq Lhx2 KO
log2(FC) E14 RNA-Seq Lhx2 KO
Casz1
Foxn4
Hes5
Sox2
Meis1
Neurod4
Notch1
Tox2
Klf4
Nfib
Nfia
Foxp4
Foxa1
F
log2(FC) P2/E14 GFP+ RNA-Seq
log2(FC) P2 ATAC-Seq Lhx2 KO
Neurod1
Lmo4
Pbx1
Rax
Hey2
Ndnf
Car2
Dlx2
Nr2e3
Atoh8
Lhx2 dependent, P2/E14 GFP+ DEG
P2/E14 GFP+ DEG
Lhx2 dependent, non DEG
G
Putative Pioneer
Sox8/9
Klf4
Zfx
Smad3
Gabpa
Sox2
NF-I
Etv5
Lhx2
Hes5
Ascl1
E14 Ctrl
E14 Lhx2 KO
P2 Ctrl
P2 Lhx2 KO
ATAC-Seq footprints

Fig.5


A
Lhx2 E14
Lhx2 P2
ATAC-Seq signal
Ctrl
Lhx2 cKO
Distance to motif center
B
Sox2 E14
Sox2 P2
C
NF-I E14
NF-I P2
D
Klf4/9 E14
Klf4/9 P2
E
Ascl1 E14
Ascl1 P2
F
Hes5 E14
Hes5 P2
G
Nucleosome occupancy
H
I
J
K
L