

1 **Title**

2 **Heterarchy of Transcription Factors Driving Basal and Luminal Cell**

3 **Phenotypes in Human Urothelium**

4

5 **Running Title**

6 Drivers of Human Urothelial Phenotype

7

8 **Authors & Affiliations**

9 Carl Fishwick<sup>1,§</sup>, Janet Higgins<sup>2</sup>, Lawrence Percival-Alwyn<sup>2</sup>, Arianna  
10 Hustler<sup>1</sup>, Joanna Pearson<sup>1</sup>, Sarah Bastkowski<sup>2</sup>, Simon Moxon<sup>2,†</sup>,  
11 David Swarbreck<sup>2</sup>, Chris D. Greenman<sup>3</sup> and Jennifer Southgate<sup>1</sup>.

12

13 <sup>1</sup>Jack Birch Unit for Molecular Carcinogenesis, Department of  
14 Biology, University of York, York YO10 5DD, United Kingdom,  
15 <sup>2</sup>Earlham Institute, Norwich Research Park, Norwich NR4 7UH,  
16 United Kingdom and <sup>3</sup>School of Computing Sciences, University of  
17 East Anglia, Norwich NR4 7TJ, United Kingdom. <sup>§</sup>Current address:  
18 Wellcome Trust Sanger Institute, Wellcome Trust Genome  
19 Campus, Hinxton, Cambridge, CB10 1SA, UK. <sup>†</sup>Current address:  
20 School of Biological Sciences, University of East Anglia, Norwich  
21 Research Park, NR4 7TJ, UK.

22 Corresponding Author: Jennifer.southgate@york.ac.uk

23 **Abstract**

24 Cell differentiation is effected by complex networks of transcription  
25 factors that co-ordinate re-organisation of the chromatin landscape.  
26 The hierarchies of these relationships can be difficult to dissect.  
27 During *in vitro* differentiation of normal human uro-epithelial cells,  
28 formaldehyde-assisted isolation of regulatory elements (FAIRE-seq)  
29 and RNA-seq were used to identify alterations in chromatin  
30 accessibility and gene expression changes following activation of the  
31 nuclear receptor PPARG as a differentiation-initiating event.

32 Regions of chromatin identified by FAIRE-seq as having altered  
33 accessibility during differentiation were found to be enriched with  
34 sequence-specific binding motifs for transcription factors predicted  
35 to be involved in driving basal and differentiated urothelial cell  
36 phenotypes, including FOXA1, P63, GRHL2, CTCF and GATA3. In  
37 addition, co-occurrence of GATA3 motifs was observed within sub-  
38 sets of differentiation-specific peaks containing P63 or FOXA1 after  
39 induction of differentiation.

40 Changes in abundance of GRHL2, GATA3, and P63 were observed in  
41 immunoblots of chromatin-enriched extracts. Transient siRNA  
42 knockdown of P63 revealed that P63 favoured a basal-like  
43 phenotype by inhibiting differentiation and promoting expression of  
44 basal marker genes. GATA3 siRNA prevented differentiation-  
45 associated downregulation of P63 protein and transcript, and  
46 demonstrated positive feedback of GATA3 on PPARG transcript, but  
47 showed no effect on FOXA1 transcript or protein expression. This

48 approach indicates that as a transcriptionally-regulated programme,  
49 urothelial differentiation operates as a heterarchy wherein GATA3 is  
50 able to co-operate with FOXA1 to drive expression of luminal marker  
51 genes, but that P63 has potential to transrepress expression of the  
52 same genes.

53 **Introduction**

54 The nuclear receptor peroxisome proliferator-activated receptor  
55 gamma (PPARG) is widely known as an essential and sufficient driver  
56 of adipogenesis (1, 2), but it also plays roles in M1 to M2  
57 polarisation of macrophages (3) and differentiation of human  
58 urothelial cells of the bladder and associated urinary tract (4-6).  
59 When grown *in vitro* in the absence of serum or other nuclear  
60 receptor signalling, non-immortalised normal human urothelial  
61 (NHU) cells acquire a proliferative, autocrine epidermal growth-  
62 factor receptor (EGFR)-regulated squamous cell phenotype (7, 8).  
63 RNA microarray studies of NHU cell cultures have shown that when  
64 downstream EGFR signalling is blocked, exogenous ligand-activation  
65 of PPARG induces expression of intermediary transcription factors  
66 required for specifying the differentiated urothelial cell phenotype,  
67 including forkhead box A1 (FOXA1), interferon regulatory factor 1  
68 (IRF1), GATA binding protein 3 (GATA3) and E74 like ETS  
69 transcription factor 3 (ELF3) (9, 10). Of these, FOXA1 and GATA3 are  
70 recognised as pioneer factors capable of driving changes in  
71 chromatin organisation and accessibility (11). In urothelial  
72 carcinoma, FOXA1 and GATA3 have been associated with  
73 differentiation status (12, 13) and 8% of tumours were found to  
74 carry ELF3 mutations (14). Mouse studies have identified other  
75 transcription factors as determinants of urothelial specification,  
76 including Grainyhead-like transcription factor 3 (Grhl3) (15),  
77 Kruppel-like factor (Klf5) (16) and Gata4 and Gata6 (17), but it  
78 remains unclear what role these factors play in human urothelium.

79 Formaldehyde-assisted isolation of regulatory elements coupled  
80 with next generation sequencing (FAIRE-seq) (18) exploits the  
81 propensity of nucleosome-depleted DNA, or “open” chromatin, to  
82 shear from adjacent nucleosomes during sonication of nuclear  
83 material from formaldehyde-fixed cells. Isolating this sheared DNA  
84 from nucleosomal DNA by phase separation enables  
85 characterisation of the relative extent of chromatin accessibility in a  
86 genome-wide manner. As transcription factors bind dynamically to  
87 nucleosome-depleted regions, motif matching within open  
88 chromatin, as identified by FAIRE, can be used to classify  
89 transcription factors that actively associate with chromatin and  
90 define cell phenotype (19-23). FAIRE identifies a complementary but  
91 partially distinct set of putative enhancer regions outside of gene  
92 promoters, as compared to DNase-seq (19) which uses DNaseI  
93 enzyme to cleave regions of open chromatin. FAIRE-seq DNA has  
94 been shown to be enriched relative to DNase-seq for potential  
95 FOXA1 binding sites, which is known to contribute to urothelial  
96 differentiation (9), and chromatin associated histone H3  
97 monomethylated on lysine 4 (H3K4me1), which is associated with  
98 genomic enhancers specific to cell type.

99 To obtain a genome-wide picture of the transcriptional drivers of  
100 different urothelial cell phenotypes, RNA-seq and FAIRE-seq were  
101 performed on serially-propagated NHU cell cultures from three  
102 independent donors at 24 h and 144 h time-points after concurrent  
103 EGFR-blockade and PPARG-activation to induce differentiation (4),  
104 alongside time-matched non-differentiated vehicle controls. Open

105 chromatin regions differentially-enriched between treated and  
106 control libraries were searched for matches to known sequence-  
107 specific transcription factor binding motifs, both on a genome-wide  
108 basis and proximal to differentially-expressed genes. Selected  
109 candidate transcriptional regulators were validated as modulators of  
110 urothelial differentiation using immunoblots of chromatin-enriched  
111 extracts and siRNA knockdown to investigate effects on urothelial  
112 phenotype.

113

## 114 **Results**

### 115 *Differentially-Expressed Genes and FAIRE-seq Peak Genomic*

#### 116 *Distribution*

117 Results obtained from the analysis of RNA-seq data identified 559  
118 and 463 genes that were upregulated, and 467 and 158 genes that  
119 were downregulated in differentiation-induced cells relative to time-  
120 matched controls at the 24 h and 144 h time-points (FDR<0.1),  
121 respectively (Supplementary Tables 1A and 1B). Genes upregulated  
122 at both time-points included the urothelium-restricted  
123 differentiation markers uroplakin 1A (UPK1A) and UPK2 (24-26).  
124 Gene ontology analysis, performed using the GOrilla tool (27),  
125 showed that the 122 genes upregulated at both time-points  
126 included genes involved in lipid metabolism ( $p=1.16 \times 10^{-5}$ ) and water  
127 homeostasis ( $p=8.09 \times 10^{-5}$ ) (Supplementary Table 2), with the latter  
128 likely reflecting the role of urothelium as a barrier to urinary solutes.

129 Peak calling using the MACS algorithm on FAIRE-seq data pooled for  
130 the three donor cell lines gave >66,000 total peaks rising to >71,000  
131 at 144 h, with a near equal distribution between proportions of  
132 distinct (control or differentiated) and overlapping peaks at each  
133 time-point (Figures 1A, B). Consistent with other investigations into  
134 the relationship between DNA enriched by FAIRE and gene  
135 expression (19, 20), when genes were split into quartiles based on  
136 normalised RNA-seq read counts (Figure 1C and Supplementary  
137 Table 3), a greater proportion of nearest-neighbour genes to FAIRE  
138 peaks had reads per kilobase per mapped million (RPKM) values  
139 above zero as compared to total genes (Figure 1D). In addition, most  
140 FAIRE peaks were intronic or intergenic, and a slight increase in the  
141 proportion of peaks associated with promoters was noted in  
142 differentiation-induced cells at both time points (Figure 1E and  
143 Supplementary Table 3).

144

#### 145 *Transcription Factor Motifs Enriched in FAIRE Peaks*

146 To uncover transcription factors driving cell phenotype in  
147 differentiated and non-differentiated urothelial cells, sequence-  
148 specific transcription factor binding motifs enriched in non-  
149 overlapping FAIRE peaks at each time-point were identified using  
150 the motif discovery tool HOMER (28). Motif searching was  
151 conducted using control-specific peak sets as the background for the  
152 differentiation-specific peak set, and vice-versa.

153 Previous transcription factor motif matching studies using open-  
154 chromatin isolation techniques have observed that particular motifs  
155 tend to be enriched at sites distal to genes (29), and that within  
156 promoter regions, transcription start sites (TSS) have fewer  
157 differences in transcription factor motifs than the rest of the  
158 genome (20). As such, FAIRE peaks in TSS promoter regions (-1kb to  
159 +100bp) were excluded from all analyses. All enrichment  
160 comparisons were performed on regions of open chromatin present  
161 either only in the control or the differentiated libraries at each time  
162 point. To highlight any differences between motifs enriched  
163 proximal to genes and those found across the genome, control-  
164 specific and differentiation-specific FAIRE peaks were compared as  
165 either genome-wide groups, or analysis was restricted to those  
166 located within  $\pm$  25 kb of the TSS of differentially-regulated genes.  
167 Motifs matched by HOMER were filtered for those which occurred in  
168 at least  $\geq$ 1.25 fold of the total percentage of regions in the target set  
169 as compared to the background set, in order to focus on motifs  
170 significantly enriched in each experimental situation (20, 30). This  
171 approach identified divergent groups of transcription factor motifs  
172 across the different regions, with each group containing matches to  
173 motifs from both previously described urothelium-associated factors  
174 and others not previously associated with urothelium (Figure 2 and  
175 Supplementary Tables 4-12). *De novo* motif analysis was less  
176 successful than matching to known motifs, as most matches that  
177 were not similar to those found in the HOMER database were in low  
178 percentages of peaks (data not shown).



179 Motifs with the highest fold-change in abundance in peaks specific  
180 to control libraries and around downregulated genes at 24 h were  
181 dominated by cell cycle-associated transcription factors such as ETS-  
182 family factors, JUN-AP1, EGR1 and EGR2, and a motif associated  
183 with combined binding of the OCT4-SOX2-TCF-NANOG pluripotency  
184 factors in mouse embryonic stem cells (31). OCT4 transcripts are  
185 expressed by NHU cells, but the pluripotency-associated isoform  
186 OCT4A is not (32). P63, a transcription factor associated with a non-  
187 differentiated “basal-like” urothelial cell phenotype in normal cells  
188 and carcinoma (33-37), was enriched both proximal to  
189 downregulated genes and across the genome at 144 h, whereas  
190 STAT6 and ETS motifs were specifically associated with peaks  $\pm 25$  kb  
191 of downregulated genes at this time-point.

192 Motifs from urothelial differentiation-associated transcription  
193 factors FOXA1 (9), GATA3 (10, 12) and PPARG (4) were enriched in  
194 differentiation-specific FAIRE peaks within  $\pm 25$  kb of the TSS of  
195 genes with expression upregulated during differentiation. PPARG  
196 motifs were only enriched around genes upregulated at 24 h, in  
197 agreement with observations that it drives early events during *in*  
198 *vitro* urothelial differentiation upstream of FOXA1 (9), motifs from  
199 which were matched at 144 h. GATA3, CEBPB and GRHL2 motifs  
200 were enriched around upregulated genes at both time points.  
201 GRHL2 has been implicated in regulation of tight junction complex  
202 genes, which are central to barrier formation in several epithelia  
203 (38), including urothelium (6, 39), whereas the closely-related  
204 GRHL3 has been associated with urothelial differentiation in the

205 mouse (15). CEBPB plays a key role in orchestrating CEBPA and  
206 PPARG expression during adipogenesis (2). CEBPB has no known role  
207 in normal human urothelial biology, although other groups have  
208 shown the CEBPB motif to be enriched in promoters of urothelial  
209 carcinoma gene sets (40), and it has been associated with urothelial  
210 differentiation in mouse (41). ELF5 and ELF1 motifs were enriched in  
211 regions proximal to upregulated and downregulated genes at 144 h,  
212 respectively. Although neither of these has been previously  
213 associated with urothelial biology, the closely related ELF3, whose  
214 motif is not in the HOMER database used here, is a driver of  
215 differentiation (10).

216 Across the genome, in differentiation-induced cells, motifs from the  
217 known urothelium-associated transcription factor IRF1 (9) and the  
218 closely related motif for IRF2 were enriched at 24 h, as were those  
219 from CTCF at both time-points. As none of these motifs were  
220 enriched proximal to differentially-regulated genes, these  
221 observations agree with previous studies which showed CTCF and  
222 IRF1 preferentially bind to regions distal to expressed genes (29).

223

#### 224 *Co-occurrence of Transcription Factor Motifs in Open Chromatin*

225 Lineage-determining transcription factors have been observed to  
226 bind in regions proximal to one another during differentiation (28).  
227 Pioneer factors such as FOXA1, which can open repressed regions of  
228 chromatin, often bind proximally to differentiation-inducing nuclear

229 receptors (42-44). To determine if there was co-occurrence of  
230 differentiation-associated transcription factor motifs within FAIRE-  
231 seq peaks, P63 and FOXA1 motif-containing open chromatin regions  
232 specific to control and differentiated cells at each time point were  
233 searched separately for enriched motifs using the same approach as  
234 for the genome-wide investigation. P63 and FOXA1 containing peaks  
235 were enriched with motifs which overlapped the overall set of  
236 peaks, but with significant differences (Supplementary Figure 2A and  
237 2B, and Supplementary Tables 13-20).

238 Motifs co-occurring within P63 and FOXA1 containing peaks were  
239 largely distinct from one another, but with notable exceptions such  
240 as GATA3, GRHL2, P63, and IRF motifs which co-occurred with both  
241 FOXA1 and P63 in differentiation-specific peaks (Supplementary  
242 Figure 3). Interestingly, OCT2, OCT4, and NF1:FOXA1 motifs were  
243 enriched in all FOXA1-containing control and differentiation-specific  
244 peak sets.

245

#### 246 *Chromatin Binding of Transcription Factors with Enriched Motifs*

247 To determine whether transcription factors with enriched motifs  
248 and other putative urothelial phenotype orchestrators reported in  
249 the literature were enriched in urothelial chromatin, immunoblots  
250 of chromatin extracts were generated using urothelial cell cultures  
251 from independent lines. PPAR $\gamma$ , FOXA1, GRHL2 and GATA3 were  
252 enriched in chromatin extracted from differentiated cell cultures,

253 whereas basal-associated P63 was more abundant in non-  
254 differentiated cultures (Figure 3). CTCF and GRHL3 had similar  
255 abundance on chromatin from control and differentiated cultures.  
256 ELF5 and ELF1 detection was not possible due to poor antibody  
257 specificity, but ELF3 was observed to be associated with chromatin  
258 from differentiated cells.

259

#### 260 *Differentiation-Associated Transcription Factors in Native*

##### 261 *Urothelium*

262 To determine if transcription factors with motifs matched to the  
263 non-differentiated or differentiated NHU cell phenotypes were  
264 expressed by normal urothelium *in situ*, immunohistochemistry was  
265 performed on human urothelial tissue sections (Figure 4). P63  
266 demonstrated a basal-intermediate cell distribution, with markedly  
267 reduced labelling of the most differentiated superficial cells. PPARG,  
268 CTCF, GATA3, GRHL2 and FOXA1 were observed to be nuclear in all  
269 layers of the urothelium, with GRHL2 and FOXA1 showing  
270 particularly intense labelling of the most differentiated superficial  
271 cell layer.

272

#### 273 *siRNA Knockdown of P63 and GATA3*

274 To further ascertain whether chromatin-associated proteins  
275 identified by FAIRE played a role in the differentiation of urothelial  
276 cells, the effects of siRNA knockdown of P63 and GATA3 on

277 expression of urothelial differentiation markers was carried out 48 h  
278 after transfection with siRNA in conjunction with differentiation or  
279 control treatment in independent NHU cell lines. In non-  
280 differentiated cells, expression of P63 protein was reduced  $\geq 2$  fold in  
281 all donors following P63 siRNA knockdown, and was reduced further  
282 following induction of differentiation (Figures 5A and D,  
283 Supplementary Figure 4). Expression of cytokeratin 13 (KRT13),  
284 which is expressed by basal and intermediate cell layers of normal  
285 human urothelium *in situ* and provides an objective marker of the  
286 switch from the basal-like squamous to a urothelial transitional  
287 epithelial differentiation programme (5), was increased following  
288 knockdown of P63 (siRNA 1) in all donors in both non-differentiated  
289 and PPARG-activated conditions, although statistical significance  
290 was not reached due to a large variation in the fold increase  
291 between different donor cell lines (Figure 5A and 5D and  
292 supplementary Figure 4). GATA3 and FOXA1 protein (measured in  
293 two NHU cell lines) increased  $\sim 2$  fold in cells treated with P63 siRNA  
294 in both non-differentiated and differentiated states (Figure 5B-D and  
295 supplementary Figure 4).

296 At the transcript level, P63 siRNA stimulated expression of KRT13  
297 and transcription factors PPARG and GATA3 in non-differentiated  
298 cells, and further increased expression of KRT13, PPARG, GATA3,  
299 FOXA1 and UPK2 transcripts following induction of differentiation  
300 (Figure 5E).

301 P63 is a key driver of genes associated with basal-type urothelial  
302 carcinomas (33, 36, 37, 45). To further investigate these links, lists of  
303 genes proximal to P63 containing motifs at the 24 h time point that  
304 overlapped genes observed to be differentially-regulated in a P63  
305 knockdown model in urothelial carcinoma cell lines (36), were  
306 compared (Supplementary Tables 21-25). Of the genes which  
307 overlapped between the P63-containing FAIRE peaks and P63  
308 knockdown in carcinoma-derived cell lines, F3, HBEGF, IGFBP3 and  
309 IL1B were further investigated by RTqPCR in P63 siRNA-treated NHU  
310 cells. In RNA-seq and during differentiation at 24 h, F3 and HBEGF  
311 were significantly downregulated, whereas IGFBP3 was upregulated  
312 (Supplementary Table 1A). Only IGFBP3 was significantly  
313 upregulated at 144 h (Supplementary Table 1B). P63 siRNA  
314 downregulated HBEGF and IL1B in the absence of differentiation  
315 inducing signals, but this effect was not observed in differentiation-  
316 induced cells for either gene (Figure 5F). IGFBP3 was strikingly  
317 upregulated in P63 siRNA-treated cells without differentiation, but  
318 only marginally upregulated in P63-siRNA cells induced to  
319 differentiate. Tissue factor F3 expression was not significantly  
320 altered by P63 siRNA in undifferentiated cells, but had weakly  
321 significantly increased expression when cells were differentiated in  
322 the presence of P63 siRNA.

323 GATA3 siRNA achieved a 1.7-7.6 fold reduction in GATA3 protein  
324 expression in differentiation-induced NHU cells, with GATA3 siRNA 2  
325 effectively abrogating the differentiation-induced increase in KRT13  
326 protein expression (Figure 6A and 6B and Supplementary Figure 5).

327 P63 protein expression was significantly upregulated in the presence  
328 of GATA3 siRNA, whereas FOXA1 protein expression was not  
329 affected.

330 GATA3 siRNA significantly attenuated transcript expression of  
331 GATA3 and the differentiation marker UPK2 (Figure 6C). KRT13  
332 transcript was only reduced significantly by GATA3 siRNA 2, as with  
333 the protein. P63 showed increases in transcript and protein  
334 expression with both GATA3 siRNA oligonucleotides. Neither GATA3  
335 siRNA sequence had an effect on FOXA1 transcript abundance and  
336 only siRNA 2 showed a small inhibitory effect on PPARG transcript  
337 expression

338

### 339 **Discussion**

340 By comparing transcription factor binding motifs matched within  
341 open chromatin regions in normal human urothelial cells in non-  
342 differentiated versus differentiated states, this study provides new  
343 insight into the identity and operational relationships between  
344 transcriptional drivers of urothelial cell phenotype. Of major  
345 significance, P63 drives the non-differentiated squamous phenotype  
346 subsumed by normal human urothelial cells maintained in serum-  
347 free culture conditions in absence of nuclear receptor signalling.  
348 Experimental knockdown revealed that P63 maintains this primitive  
349 or “basal-like” phenotype at least in part by inhibiting expression of  
350 transitional epithelial lineage genes including KRT13 and PPARG.

351

352 P63 plays an essential role in epithelial tissues during development,  
353 where its absence causes severe dysgenesis of epithelial tissues, as  
354 described in mouse epidermis (46). Changes in expression and  
355 somatic mutations of P63 have been associated with clinically-  
356 relevant subtypes of bladder cancer, with P63 identified as a driver  
357 of the basal-like cell phenotype in urothelial carcinoma (36). These  
358 authors showed that knockdown of P63 in the established bladder  
359 cancer-derived UM-UC14 cell line affected expression of PPARG-  
360 influenced genes, including HBEGF, IGFBP3 and IL1B (36). Here,  
361 these same genes were differentially affected by siRNA knockdown  
362 of P63 in NHU cells, implying usage of the same gene networks by  
363 normal and cancer cells.

364 In urothelium, PPARG has been identified as a nuclear receptor  
365 whose activation mediates the transition from squamous to a  
366 differentiated transitional (urothelial) phenotype. This involves a  
367 major shift in gene expression, implying a change in genomic  
368 organisation to reflect the transcriptional landscape of urothelium.  
369 We have previously identified a network of PPARG-regulated  
370 intermediary transcription factors that mediate the differentiated  
371 urothelial programme, although inter-relationships within the  
372 network have yet to be established. In other tissues, such as breast,  
373 a role has been identified for the so-called pioneer factors FOXA  
374 and GATA in defining the tissue-specific genomic organisation.  
375 GATA3 and FOXA1 have been shown to act co-operatively in



376 promoting ESR1-driven transcription in MCF7 cells, with GATA3 lying  
377 upstream of FOXA1(44). In the current study, GATA3 siRNA in  
378 combination with PPARG stimulation prevented downregulation of  
379 P63 and attenuated expression of intermediate to late  
380 differentiation markers, but did not alter FOXA1 expression. As  
381 FOXA1, P63 and GATA3 motifs were all co-enriched within the same  
382 open chromatin associated specifically with differentiation, this  
383 establishes a basis for a model of the interaction of all three factors  
384 in determining urothelial phenotype wherein P63 outcompetes  
385 FOXA1 for chromatin binding sites in the absence of GATA3. The  
386 results from modulating GATA3 expression point to the existence of  
387 a heterarchical relationship between differentiation drivers, in  
388 which transcription factors such as GATA3 are capable of influencing  
389 the expression of phenotypic drivers such as P63 independently of  
390 other key determining intermediary transcription factors in the  
391 network, including FOXA1.

392 The motif-matching performed here identified transcription factors  
393 not previously associated with urothelial differentiation, including  
394 CTCF. CTCF was not enriched at the protein level in chromatin  
395 extracts after induction of differentiation, most probably because  
396 CTCF is a constitutive chromatin-associated protein which facilitates  
397 looping between promoters and enhancers (47-51). The results in  
398 this study add to the weight of evidence that CTCF binding, although  
399 widespread and well-conserved in many genomic regions (47-51),  
400 shows tissue-specific genome binding activity around genes that  
401 specify cell type-specific phenotypes.

402 Our initial analysis of differentially-expressed gene transcripts  
403 indicated a potential role for GRHL3 in differentiation of human  
404 urothelium. However, no differentiation-associated changes in  
405 GRHL3 protein abundance or localisation were seen during  
406 differentiation, whereas the constitutively expressed GRHL2 gene  
407 showed increased protein abundance and relocation onto the  
408 chromatin of differentiating cells. Taken with the nuclear localisation  
409 of GRHL2 in situ, these data implicate GRHL2 as the more important  
410 player and further illustrate that not all differentiation-associated  
411 events are transcriptionally-regulated. GRHL2 has been observed to  
412 be expressed by human urothelium in another recent study (52) and  
413 is known to reside within a genomic region which is commonly  
414 amplified in aggressive urothelial carcinoma (53).

415 Another novel factor was KLF5, which was shown to be expressed  
416 constitutively by RNA-seq and moderately, albeit not statistically  
417 significantly, upregulated in response to differentiation at both time  
418 points investigated. Klf5 is reported to be upstream of Pparg and  
419 Grhl3 in mouse urothelial development (16), suggesting it may  
420 function in early urothelial specification and not be directly  
421 associated with regulating genes associated with mature  
422 differentiation stages. Klf5 and Gata4 have been associated with  
423 urothelial differentiation in mouse (16, 17). However, GATA4 was  
424 not detected in RNA-seq data in the current study, where GATA3  
425 transcript was the most highly expressed GATA gene family member  
426 detected and in addition, was the only GATA member to be  
427 upregulated upon differentiation and associated with post-

428 differentiation chromatin. These data implicate GATA3 rather than  
429 GATA4 in the differentiation of human urothelium and again this is  
430 supported in vivo, at least indirectly by immunohistochemical  
431 studies in situ.

432 **Methods**

433 *In Vitro Growth and Differentiation of Normal Human Urothelial*

434 *Cells*

435 Normal human urothelial (NHU) cells were maintained as finite,  
436 serially-passaged cell lines, as described previously (54). Cultures  
437 were sub-cultured by trypsinisation and maintained in Keratinocyte  
438 Serum Free Medium containing bovine pituitary extract and  
439 epidermal growth factor (Gibco) and further supplemented with 30  
440 ng/ml cholera toxin (Sigma). Differentiation was induced in just-  
441 confluent cell cultures using 1 $\mu$ M troglitazone as PPAR $\gamma$  ligand with  
442 concurrent 1  $\mu$ M PD153035 to block EGFR activation (4). Non-  
443 differentiated vehicle control (0.1% DMSO) cultures were  
444 maintained in parallel and used at the same time points (24 and 144  
445 hours).

446

447 *RNA-seq Sample and Library Preparation*

448 Cell monolayers were solubilised in Trizol (Life Technologies), using  
449 the manufacturer's protocol for chloroform and isopropanol  
450 extraction, and DNA was digested using RNase-free DNase I  
451 (Ambion). Library construction was performed using TruSeq RNA  
452 Sample Prep Kit v2 (Illumina). Sequencing was performed using an  
453 Illumina HiSeq 2500 sequencer and reads aligned using RSEM (55) to  
454 the reference UCSC hg19 human genome. Differential gene  
455 expression was performed between control and differentiation-

456 induced cells at 24hr and 144hr time-points using DESeq (56). The  
457 results obtained from three independent cell lines were treated as  
458 replicates and genes with a false discovery rate (FDR) cut-off <0.1  
459 were called significant.

460

#### 461 *FAIRE-seq Sample and Library Preparation*

462 Cell monolayers were fixed in 1% formaldehyde for 10 minutes  
463 before quenching by addition of glycine to 125 mM for 5 minutes  
464 and scrape-harvesting in ice-cold PBS with added protease  
465 inhibitors. Approximately  $5 \times 10^6$  cells were lysed and sheared, and  
466 open chromatin extracted as described in the FAIRE protocol (57).

467

#### 468 *Motif Searching*

469 MACS peak-calling algorithm (58) was used to call FAIRE-enriched  
470 peaks. Non-overlapping peaks between control and differentiated  
471 samples at each time point were identified using bedtools. HOMER  
472 motif discovery software (28) was used to discover motifs over-  
473 represented in each treatment condition, using peaks uniquely  
474 present in control cells as the background when searching the  
475 differentiation-induced specific peaks, and vice versa. Motifs  
476 identified by HOMER as enriched were further filtered by fold-  
477 change as percentage enrichment above background of  $\geq 1.25$ .

478

479 *Chromatin Enrichment*

480 Cells were fixed and scrape-harvested as for FAIRE, then pelleted  
481 cells were subjected to a chromatin enrichment protocol (59) with  
482 optional RNase digestion step included.

483

484 *Antibodies*

485 Anti-FOXA1 (Santa Cruz, Catalogue # sc-101058) used at 1:250 for  
486 IHC and 1:400 for immunoblot. Anti-CTCF (Cell Signalling, Catalogue  
487 #2899) used at 1:250 for IHC and 1:1000 for immunoblot. Anti-P63  
488 (Santa Cruz Biotechnologies, Catalogue # sc-8431) used at 1:1000 for  
489 IHC and 1:500 for immunoblot. Anti-GRHL2 (Abcam, Catalogue #  
490 ab88631) used at 1:150 for IHC and 1:400 for immunoblot. Anti-  
491 PPARG (Santa Cruz, Catalogue # 7273) used at 1:2000 for IHC and  
492 1:500 for immunoblot. Anti-GATA3 (Cell signalling, Catalogue #  
493 5852) used at 1:800 for IHC and 1:200 for immunoblot. Anti-GRHL3  
494 (Abcam, Catalogue # ab57612) used at 1:500 for immunoblot. Anti  
495 KRT13 (Abnova, Catalogue # MAB1864) used at 1:1000 for  
496 immunoblot. Anti-BACT (Sigma-Aldrich, Catalogue # AC5441) used at  
497 1:250,000 for immunoblot).

498 **Acknowledgements**

499 Next-generation sequencing and library construction was delivered  
500 via the BBSRC National Capability in Genomics (BB/J010375/1) at the  
501 Earlham Institute by members of the Platforms and Pipelines Group.  
502 CF was a BBSRC-funded CASE student with GSK. All other parts of  
503 the study were funded by York Against Cancer. Jennifer Hinley is  
504 thanked for the immunohistochemistry work.

505

506 **Conflict of Interest Statement**

507 The authors confirm that there are no competing financial interests.

508

509 **Supplementary Information**

510 Supplementary information is available at Cell Death and  
511 Differentiation's website.

## REFERENCES

1. Tontonoz P, Hu E, Spiegelman BM. Stimulation of adipogenesis in fibroblasts by PPAR $\gamma$ 2, a lipid-activated transcription factor. *Cell*. 1994;79(7):1147-56.
2. Wu Z, Xie Y, Bucher NL, Farmer SR. Conditional ectopic expression of C/EBP beta in NIH-3T3 cells induces PPAR gamma and stimulates adipogenesis. *Genes Dev*. 1995;9(19):2350-63.
3. Lefterova MI, Steger DJ, Zhuo D, Qatanani M, Mullican SE, Tuteja G, et al. Cell-specific determinants of peroxisome proliferator-activated receptor gamma function in adipocytes and macrophages. *Mol Cell Biol*. 2010;30(9):2078-89.
4. Varley CL, Stahlschmidt J, Lee WC, Holder J, Diggle C, Selby PJ, et al. Role of PPARgamma and EGFR signalling in the urothelial terminal differentiation programme. *J Cell Sci*. 2004;117(Pt 10):2029-36.
5. Varley CL, Stahlschmidt J, Smith B, Stower M, Southgate J. Activation of peroxisome proliferator-activated receptor-gamma reverses squamous metaplasia and induces transitional differentiation in normal human urothelial cells. *Am J Pathol*. 2004;164(5):1789-98.
6. Varley CL, Garthwaite MAE, Cross W, Hinley J, Trejdosiewicz LK, Southgate J. PPAR -regulated tight junction development during human urothelial cytodifferentiation. *J Cell Physiol*. 2006;208(2):407.
7. Varley C, Hill G, Pellegrin S, Shaw NJ, Selby PJ, Trejdosiewicz LK, et al. Autocrine regulation of human urothelial cell proliferation and migration during regenerative responses in vitro. *Exp Cell Res*. 2005;306(1):216-29.
8. Rebouissou S, Bernard-Pierrot I, de Reyniès A, Lepage M-L, Krucker C, Chapeaublanc E, et al. EGFR as a potential therapeutic target for a subset of muscle-invasive bladder cancers presenting a basal-like phenotype. *Sci Trans Med*. 2014;6(244):244ra91-ra91.
9. Varley C, Bacon E, Holder J, Southgate J. FOXA1 and IRF-1 intermediary transcriptional regulators of PPAR -induced urothelial cytodifferentiation. *Cell Death Different*. 2008;16(1):103-14.
10. Böck M, Hinley J, Schmitt C, Wahlicht T, Kramer S, Southgate J. Identification of ELF3 as an early transcriptional regulator of human urothelium. *Dev Biol*. 2014;386(2):321-30.
11. Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell*. 2002;9(2):279-89.
12. Higgins JP, Kaygusuz G, Wang L, Montgomery K, Mason V, Zhu SX, et al. Placental S100 (S100P) and GATA3: markers for transitional epithelium and urothelial carcinoma discovered by complementary DNA microarray. *Am J Surg Pathol*. 2007;31(5):673-80.
13. DeGraff DJ, Clark PE, Cates JM, Yamashita H, Robinson VL, Yu X, et al. Loss of the Urothelial Differentiation Marker FOXA1 Is Associated with High Grade, Late Stage Bladder Cancer and Increased Tumor Proliferation. *PLoS ONE*. 2012;7(5):e36669.



14. The Cancer Genome Atlas Research N. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*. 2014;507(7492):315-22.
15. Yu Z, Mannik J, Soto A, Lin KK, Andersen B. The epidermal differentiation-associated Grainyhead gene *Get1/Grhl3* also regulates urothelial differentiation. *EMBO J*. 2009;28(13):1890-903.
16. Bell SM, Zhang L, Mendell A, Xu Y, Haitchi HM, Lessard JL, et al. Kruppel-like factor 5 is required for formation and differentiation of the bladder urothelium. *Dev Biol*. 2011;358(1):79-90.
17. Mauney JR, Ramachandran A, Yu RN, Daley GQ, Adam RM, Estrada CR. All-trans retinoic acid directs urothelial specification of murine embryonic stem cells via *GATA4/6* signaling mechanisms. *PLoS ONE*. 2010;5(7):e11513.
18. Giresi PG, Kim J, McDaniel RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res*. 2007;17(6):877-85.
19. Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, Lee B-K, et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res*. 2011;21(10):1757-67.
20. Waki H, Nakamura M, Yamauchi T, Wakabayashi K, Yu J, Hirose-Yotsuya L, et al. Global mapping of cell type-specific open chromatin by FAIRE-seq reveals the regulatory role of the NFI family in adipocyte differentiation. *PLoS Genet*. 2011;7(10):e1002311.
21. Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, et al. A map of open chromatin in human pancreatic islets. *Nat Genet*. 2010;42(3):255-9.
22. Murtha M, Strino F, Tokcaer-Keskin Z, Sumru Bayin N, Shalabi D, Xi X, et al. Comparative FAIRE-seq analysis reveals distinguishing features of the chromatin structure of ground state- and primed-pluripotent cells. *Stem Cells*. 2015;33(2):378-91.
23. Reschen ME, Gaulton KJ, Lin D, Soilleux EJ, Morris AJ, Smyth SS, et al. Lipid-Induced Epigenomic Changes in Human Macrophages Identify a Coronary Artery Disease-Associated Variant that Regulates *PPAP2B* Expression through Altered C/EBP-Beta Binding. *PLoS Gen*. 2015;11(4):e1005061-e.
24. Sun TT, Liang FX, Wu XR. Uroplakins as markers of urothelial differentiation. *Adv Exp Med Biol*. 1999;462:7.
25. Lobban ED, Smith BA, Hall GD, Harnden P, Roberts P, Selby PJ, et al. Uroplakin gene expression by normal and neoplastic human urothelium. *Am J Pathol*. 1998;153(6):1957-67.
26. Olsburgh J, Harnden P, Weeks R, Smith B, Joyce A, Hall G, et al. Uroplakin gene expression in normal human tissues and locally advanced bladder cancer. *J Pathol*. 2003;199(1):41-9.
27. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC bioinformatics*. 2009;10(1):48.
28. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple Combinations of Lineage-Determining Transcription Factors

- Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular cell*. 2010;38(4):576-89.
29. Boyle AP, Song L, Lee BK, London D, Keefe D, Birney E, et al. High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. *Genome Res*. 2011;21(3):456-64.
30. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*. 2007;448(7153):553-60.
31. Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*. 2008;133(6):1106-17.
32. Wezel F, Pearson J, Kirkwood LA, Southgate J. Differential expression of Oct4 variants and pseudogenes in normal urothelium and urothelial cancer. *Am J Pathol*. 2013;183(4):1128-36.
33. Karni-Schmidt O, Castillo-Martin M, Shen TH, Gladoun N, Domingo-Domenech J, Sanchez-Carbayo M, et al. Distinct expression profiles of p63 variants during urothelial development and bladder cancer progression. *Am J Pathol*. 2011;178(3):1350-60.
34. Feil G, Maurer S, Nagele U, Krug J, Bock C, Sievert KD, et al. Immunoreactivity of p63 in monolayered and in vitro stratified human urothelial cell cultures compared with native urothelial tissue. *Eur Urol*. 2008;53(5):1066-72.
35. Wezel F, Pearson J, Southgate J. Plasticity of in vitro-generated urothelial cells for functional tissue formation. *Tissue Eng Part A*. 2014;20(9-10):1358-68.
36. Choi W, Porten S, Kim S, Willis D, Plimack ER, Hoffman-Censits J, et al. Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer cell*. 2014;25(2):152-65.
37. Tran MN, Choi W, Wszolek MF, Navai N, Lee I-LC, Nitti G, et al. The p63 Protein Isoform  $\Delta Np63\alpha$  Inhibits Epithelial-Mesenchymal Transition in Human Bladder Cancer Cells ROLE OF MIR-205. *J Biol Chem*. 2013;288(5):3275-88.
38. Werth M, Walentin K, Aue A, Schonheit J, Wuebken A, Podewski N, et al. The transcription factor grainyhead-like 2 regulates the molecular composition of the epithelial apical junctional complex. *Development*. 2010;137(22):3835-45.
39. Smith NJ, Hinley J, Varley CL, Eardley I, Trejdosiewicz LK, Southgate J. The human urothelial tight junction: claudin 3 and the ZO-1alpha switch. *Bladder*. 2015;2(1):e9.
40. Eriksson P, Aine M, Veerla S, Liedberg F, Sjö Dahl G, Höglund M. Molecular subtypes of urothelial carcinoma are defined by specific gene regulatory systems. *BMC Med Gen*. 2015;8(1):1.
41. Dozmorov M, Stone R, 2nd, Clifford JL, Sabichi AL, Engles CD, Hauser PJ, et al. System level changes in gene expression in maturing bladder mucosa. *J Urol*. 2011;185(5):1952-8.

42. Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoute J, et al. Genome-wide analysis of estrogen receptor binding sites. *Nat Genet.* 2006;38(11):1289-97.
43. Lupien M, Eeckhoute J, Meyer C, Wang Q, Zhang Y, Li W, et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell.* 2008;132(6):958-70.
44. Theodorou V, Stark R, Menon S, Carroll JS. GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome Res.* 2013;23(1):12-22.
45. Choi W, Shah JB, Tran M, Svatek R, Marquis L, Lee I-L, et al. p63 expression defines a lethal subset of muscle-invasive bladder cancers. *PLoS ONE.* 2012;7(1):e30206.
46. Mills AA, Zheng B, Wang X-J, Vogel H, Roop DR, Bradley A. p63 is a p53 homologue required for limb and epidermal morphogenesis. *Nature.* 1999;398(6729):708-13.
47. Ouboussad L, Kreuz S, Lefevre PF. CTCF depletion alters chromatin structure and transcription of myeloid-specific factors. *J Mol Cell Biol.* 2013;5(5):308-22.
48. Dubois-Chevalier J, Oger F, Dehondt H, Firmin FF, Gheeraert C, Staels B, et al. A dynamic CTCF chromatin binding landscape promotes DNA hydroxymethylation and transcriptional induction of adipocyte differentiation. *Nucleic Acids Res.* 2014;42(17):10943-59.
49. Heath H, Ribeiro de Almeida C, Sleutels F, Dingjan G, van de Nobelen S, Jonkers I, et al. CTCF regulates cell cycle progression of alphabeta T cells in the thymus. *EMBO J.* 2008;27(21):2839-50.
50. Soshnikova N, Montavon T, Leleu M, Galjart N, Duboule D. Functional analysis of CTCF during mammalian limb development. *Dev Cell.* 2010;19(6):819-30.
51. Sanborn AL, Rao SS, Huang S-C, Durand NC, Huntley MH, Jewett AI, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Nat Acad Sci.* 2015;112(47):E6456-E65.
52. Riethdorf S, Frey S, Santjer S, Stoupiet M, Otto B, Riethdorf L, et al. Diverse expression patterns of the EMT suppressor grainyhead-like 2 (GRHL2) in normal and tumour tissues. *Int J Cancer.* 2016;138(4):949-63.
53. Hurst CD, Platt FM, Taylor CF, Knowles MA. Novel Tumor Subgroups of Urothelial Carcinoma of the Bladder Defined by Integrated Genomic Analysis. *Clin Cancer Res.* 2012;18(21):5865-77.
54. Southgate J, Hutton KA, Thomas DF, Trejdosiewicz LK. Normal human urothelial cells in vitro: proliferation and induction of stratification. *Lab Invest.* 1994;71(4):583-94.
55. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* 2011;12(1):323.
56. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010;11(10):R106.
57. Giresi PG, Lieb JD. Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). *Methods.* 2009;48(3):233-9.

58. Feng J, Liu T, Zhang Y. Using MACS to Identify Peaks from CHIP-Seq Data. *Curr Prot Bioinform.* 2011;2.14. 1-2.. .
59. Kustatscher G, Wills KL, Furlan C, Rappsilber J. Chromatin enrichment for proteomics. *Nat Protoc.* 2014;9(9):2090-9.

## Figure Captions

Figure 1. (A, B) Numbers of overlapping FAIRE peaks between control and differentiation-induced cells at 24 and 144 h. (C) Genes for the 24 h control sample were split into quartiles based on RPKM in RNA-seq data, High RPKM  $\geq 10$ , Medium RPKM  $\geq 1 < 10$ , Low RPKM  $> 0 > 1$ , Zero RPKM = 0. (D) FAIRE peaks were labelled based on the high, medium or zero expression of the nearest neighbour gene. FAIRE peaks were in this case more often near genes with expression above zero. Data representative of all time points. (E) Position of FAIRE peaks relative to annotated genomic features demonstrated that the majority of peaks were intronic or intergenic. The greatest variation between samples was seen within the proportion of peaks at promoters directly upstream of a transcription start site, with increases in the proportion of FAIRE peaks at these sites in both differentiated time-points relative to their non-differentiated controls.

Figure 2. Summary of known motifs from the HOMER database matched in FAIRE-seq peaks specific to control and differentiation-induced NHU cells. FAIRE-seq peaks from pooled donor data were compared between control and differentiation-induced cells at 24 and 144 h time-points, and peaks unique (non-overlapping) to each library were searched for known sequence motifs in HOMER to generate a genome-wide comparison for all peaks. The same comparison was performed using only peaks found within  $\pm 25$  kb of

the TSS of genes upregulated or downregulated during differentiation at the respective time-points.

Figure 3. Chromatin extracts showing bound transcription factors which changed in abundance during differentiation. Factors with motifs detected as enriched in differentiation-specific FAIRE peaks, including GRHL2, GATA3, FOXA1 and PPARG were upregulated in chromatin extracts from differentiation-induced NHU cells from two independent donors. CTCF and GRHL3 did not change in abundance with differentiation. P63 abundance was reduced after induction of differentiation. Histone H2A is included as a loading control.

Figure 4. Native human urothelium showed nuclear localisation of differentiation-associated transcription factors CTCF, FOXA1, GATA3, GRHL2 and PPARG in all stratified layers. P63 was observed predominantly in basal and intermediate cells. Occasional cells in the urothelium with condensed nuclei which do not label for most transcription factors are consistent, morphologically, with infiltrating lymphocytes.

Figure 5. Immunoblot of whole cell lysates from representative NHU cell donors showing effect of P63 siRNA on (A) P63 and KRT13, (B) FOXA1, and (C) GATA3 protein expression, with (+) and without (-) differentiation induction at 48 h. ACTB =  $\beta$  actin loading control.

FOXA1 and GATA3 were on the same membrane and normalised to the ACTB shown with FOXA1. (D) Densitometry measurements from immunoblots showing  $\log_{(2)}$  fold change of intensity in immunoblotting for three independent donors for P63 and KRT13, and two independent donors for GATA3 and FOXA1 following P63 siRNA, relative to control siRNA. Statistical test performed where material from three donors was measured was a Repeated Measures one-way ANOVA with Greenhouse-Geisser correction and Sidak's multiple comparison post-test, with p-values indicated by \* ( $P \leq 0.05$ ), \*\* ( $P \leq 0.01$ ), \*\*\* ( $P \leq 0.001$ ) and \*\*\*\* ( $P < 0.0001$ ). (E, F) RT-QPCR results from NHU cells from three independent donors showing change in abundance of RNA transcript after exposure to P63 siRNA either with or without induction of differentiation for 48 h for (E) urothelial differentiation-associated, and (F) genes associated with P63 motif containing FAIRE peaks.  $\log_{(2)}$  fold change measured relative to control siRNA with or without differentiation induction. All qPCR transcript relative abundance measurements were normalised internally to GAPDH. Statistics was performed using a two-way ANOVA with Dunnett's multiple comparison post-test, with P values indicated by \* ( $P \leq 0.05$ ), \*\* ( $P \leq 0.01$ ), \*\*\* ( $P \leq 0.001$ ) and \*\*\*\* ( $P < 0.0001$ ).

Figure 6. (A) Representative immunoblots of NHU whole cell lysate showing GATA3, KRT13, FOXA1 and P63 protein expression after differentiation induction for 48 h following transfection with GATA3 siRNA. ACTB =  $\beta$  actin loading control. KRT13, FOXA1 and GATA3

were on the same membrane and normalised to the ACTB shown, and the P63 was on a separate membrane and normalised to a separate ACTB as shown in the supplementary data. (B)

Densitometry measurements from immunoblots of three donors showing  $\log_{(2)}$  fold change in expression of GATA3, KRT13 and FOXA1 in 48 h differentiation-induced NHU cells following transfection with GATA3 siRNA relative to control siRNA. Signals for P63 and KRT13 were normalised for loading to  $\beta$ -actin (ACTB) and fold change determined relative to the equivalent control siRNA transfection results. Statistical test performed was a Repeated Measures one-way ANOVA with Greenhouse-Geisser correction and Sidak's multiple comparison post-test, with P-values indicated by \* ( $P \leq 0.05$ ), \*\* ( $P \leq 0.01$ ), \*\*\* ( $P \leq 0.001$ ) and \*\*\*\* ( $P < 0.0001$ ). (C) RT-QPCR results combined from NHU cells from three independent donors showing change in abundance of RNA transcript for P63 and differentiation-associated genes after transfection with GATA3 siRNA followed by differentiation for 48 h.  $\log_{(2)}$  fold change shown relative to control siRNA transfection with followed by 48 h differentiation. Statistics was performed using a two-way ANOVA with Dunnett's multiple comparison post-test, with P-values indicated by \* ( $P \leq 0.05$ ), \*\* ( $P \leq 0.01$ ), \*\*\* ( $P \leq 0.001$ ) and \*\*\*\* ( $P < 0.0001$ ).

### **Supplementary Figure Captions**

Supplementary Figure 1. Numbers of differentially-regulated genes by RNA-seq at 24 h and 144 h post induction of differentiation.



Supplementary Figure 2. (A) Numbers of motifs co-occurring in peaks containing FOXA1 or P63 overlapping with the overall set of peaks specific to each time point and treatment condition. (B) Numbers of overlapping peaks between P63 and FOXA1 containing peaks at each time-point.

Supplementary Figure 3. Co-occurring motifs in P63 and FOXA1 containing peaks.

Supplementary Figure 4. Immunoblot of whole cell lysates from additional NHU cell donors showing effect of P63 siRNA on (A, B) P63 and KRT13, (C) FOXA1, and GATA3 protein expression, with (+) and without (-) differentiation induction at 48 h. ACTB =  $\beta$  actin loading control. FOXA1 and GATA3 were on the same membrane and normalised to the ACTB shown with FOXA1.

Supplementary Figure 5. Additional NHU whole cell lysate showing (A) GATA3, KRT13 and FOXA1, and (B) P63 expression after differentiation induction for 48 h following transfection with GATA3 siRNA. ACTB =  $\beta$  actin loading control. KRT13, FOXA1 and GATA3 were blotted on the same membrane and normalised to the ACTB shown with KRT13. P63 was on a separate membrane and normalised to a the ACTB shown directly below. All three P63 blots are shown, including that shown in Figure 6A.

#### **Supplementary Table Captions**

Supplementary Table 1A. Genes differentially expressed between non-differentiated and differentiated cells at 24 h.

Supplementary Table 1B. Genes differentially expressed between non-differentiated and differentiated cells at 24 h.

Supplementary Table 2. Gene ontology analysis of genes upregulated by differentiation at both 24 h and 144 h using GORilla.

Supplementary Table 3. Expression quartiles for all genes and genes within 25 kb of FAIRE peaks.

Supplementary Table 4. Summary of HOMER motif results for all FAIRE peaks and FAIRE peaks within  $-/+$  25 kb of differentially expressed genes.

Supplementary Table 5. Motifs enriched in all FAIRE peaks unique to control cells at 24 h.

Supplementary Table 6. Motifs enriched in all FAIRE peaks unique to differentiated cells at 24 h.

Supplementary Table 7. Motifs enriched in all FAIRE peaks unique to control cells at 144 h.

Supplementary Table 8. Motifs enriched in all FAIRE peaks unique to differentiated cells at 144 h.

Supplementary Table 9. Motifs enriched in FAIRE peaks within 25 kb of genes downregulated after 24 h differentiation.

Supplementary Table 10. Motifs enriched in FAIRE peaks within 25 kb of genes upregulated after 24 h differentiation.

Supplementary Table 11. Motifs enriched in FAIRE peaks within 25 kb of genes downregulated after 144 h differentiation.

Supplementary Table 12. Motifs enriched in FAIRE peaks within 25 kb of genes upregulated after 144 h differentiation.

Supplementary Table 13. HOMER motif results for FOXA1-containing peaks at 24 h control.

Supplementary Table 14. HOMER motif results for FOXA1-containing peaks at 24 h differentiation.

Supplementary Table 15. HOMER motif results for FOXA1-containing peaks at 144 h control.

Supplementary Table 16. HOMER motif results for FOXA1-containing peaks at 144 h differentiation.

Supplementary Table 17. HOMER motif results for FOXA1-containing peaks at 24 h control.

Supplementary Table 18. HOMER motif results for FOXA1-containing peaks at 24 h differentiation.

Supplementary Table 19. HOMER motif results for FOXA1-containing peaks at 144 h control.

Supplementary Table 20. HOMER motif results for FOXA1-containing peaks at 144 h differentiation.

Supplementary Table 21. MACS peaks containing P63 in 24 h control peaks.

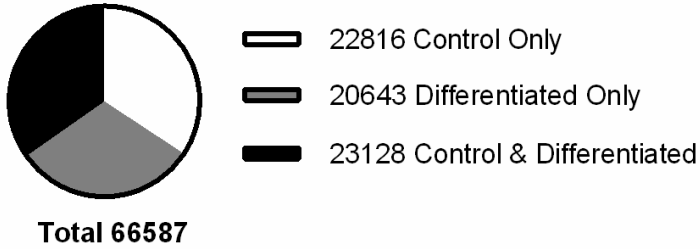
Supplementary Table 22. MACS peaks containing P63 in 24 h differentiated peaks.

Supplementary Table 23. Gene annotations for MACS peaks in 24 h control peaks.

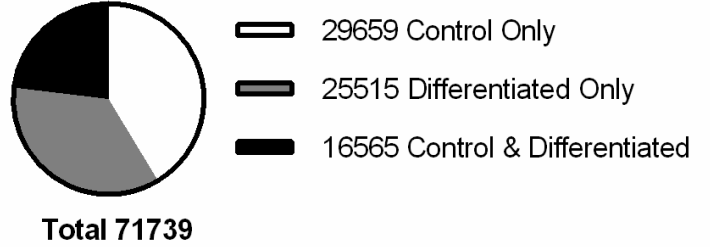
Supplementary Table 24. Gene annotations for MACS peaks in 24 h differentiated peaks.

Supplementary Table 25. Comparison of genes annotated with P63 containing peaks and genes differentially expressed after P63 knockdown in UM-UC14 cells.

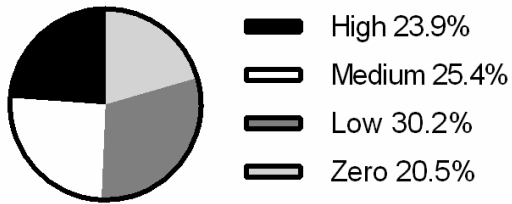
**A FAIRE Peaks Shared (24 h)**



**B FAIRE Peaks Shared (144 h)**



**C Transcript Abundance (All Genes)**

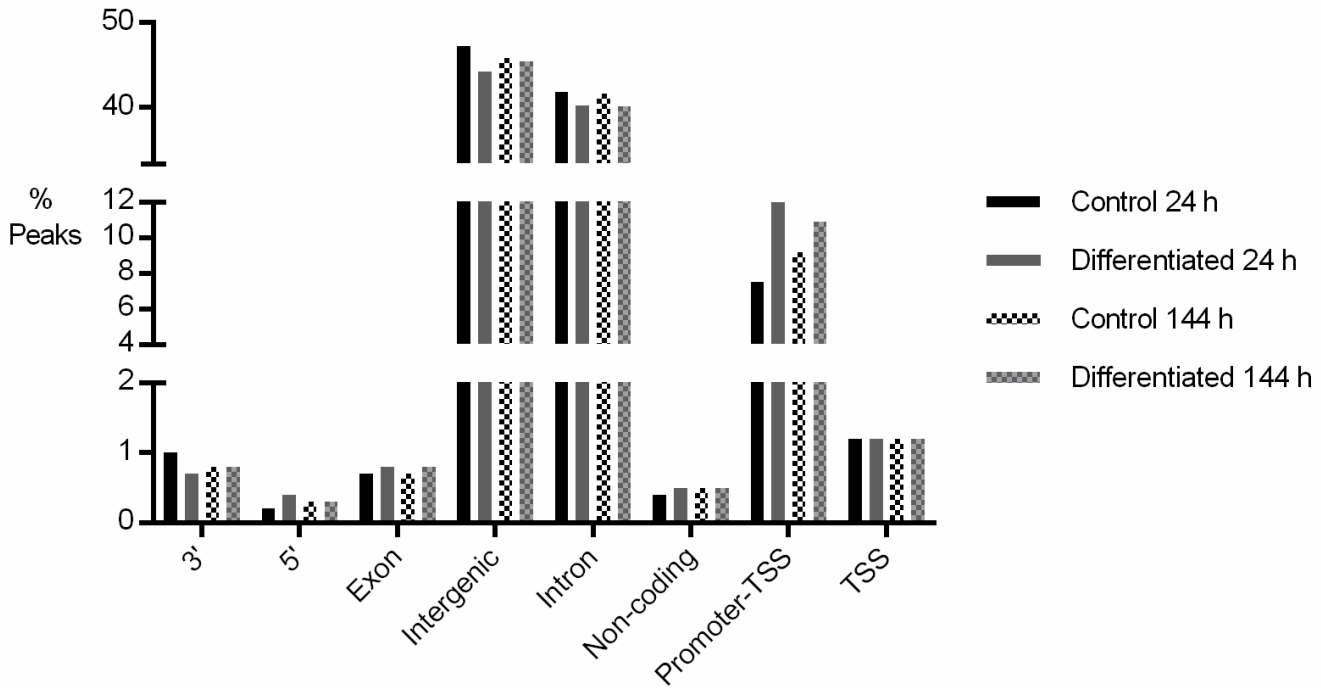


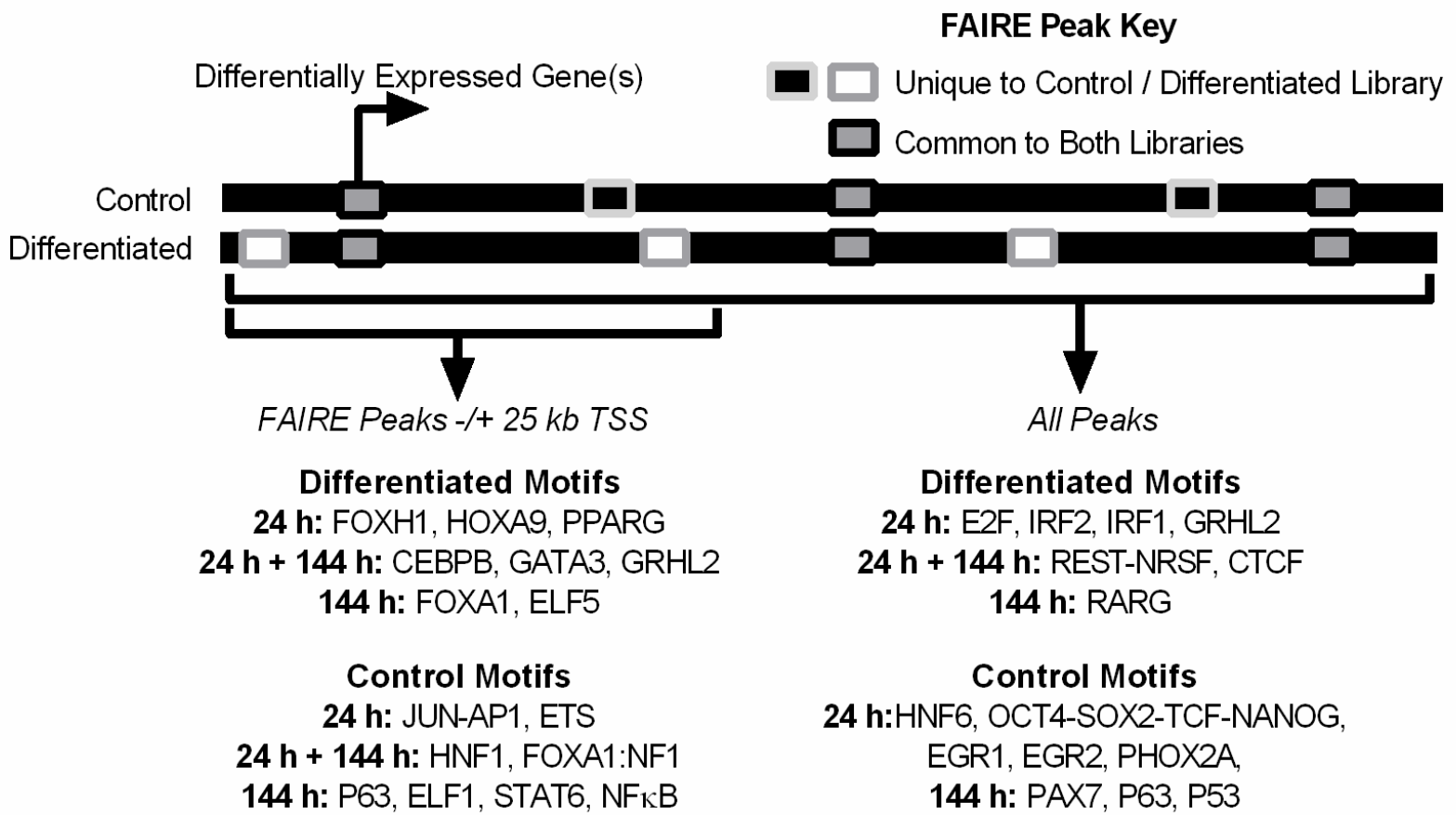
**D Expression of FAIRE Peak Nearest-Neighbour Gene**



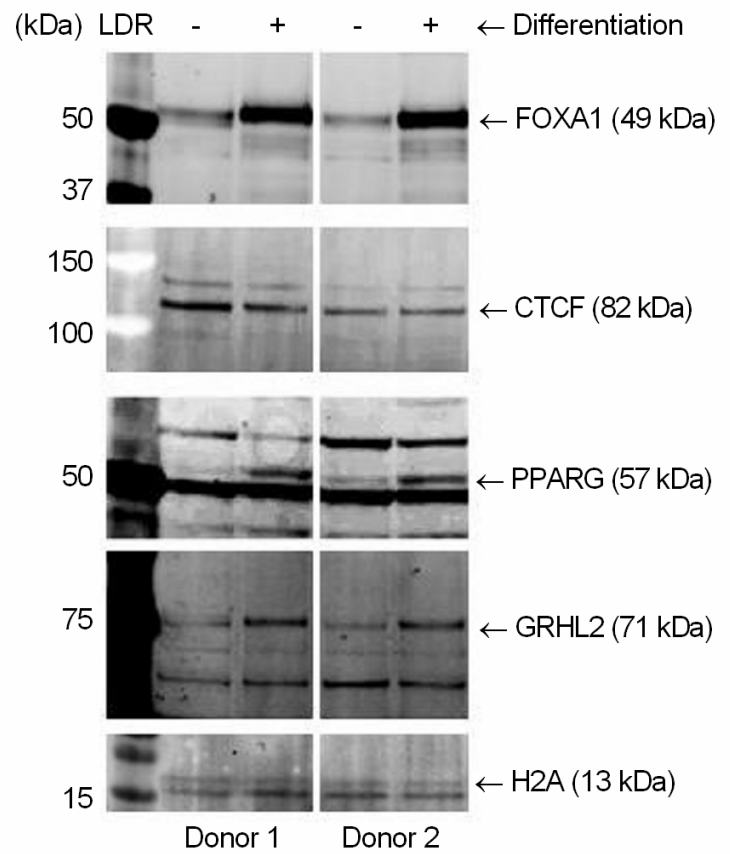
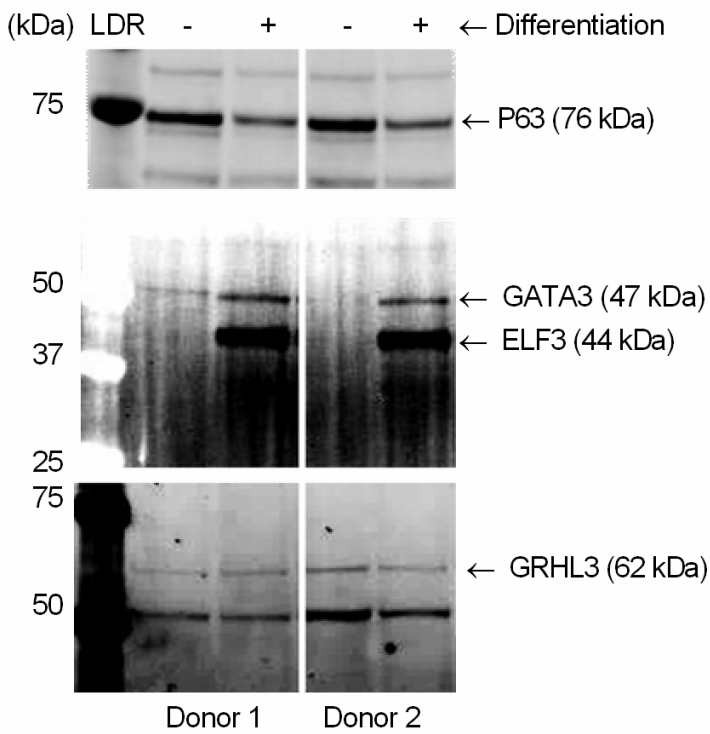
**E**

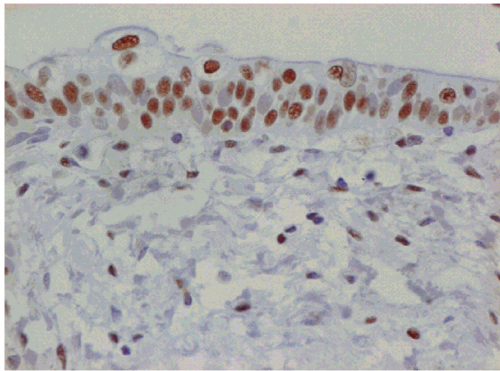
**FAIRE Peak Genomic Distribution**



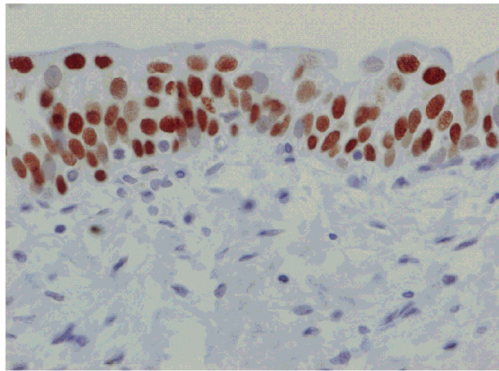


Motifs filtered for  $p < 0.05$  and Fold Change in % Occurrence Versus Background  $\geq 1.25$

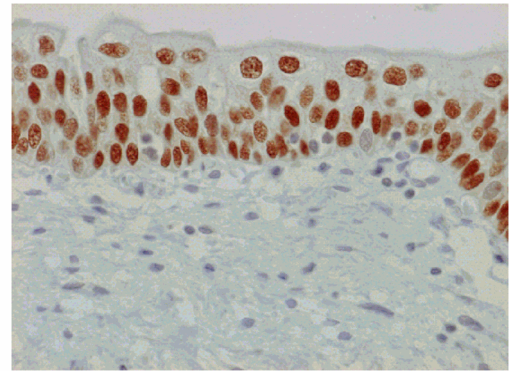




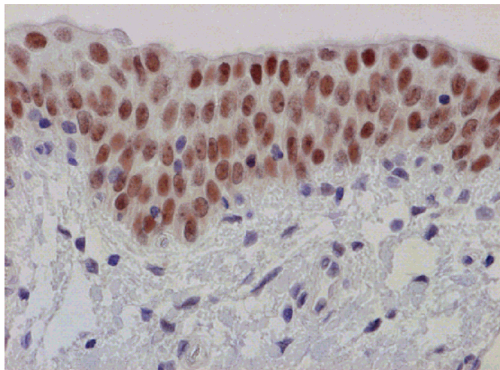
CTCF



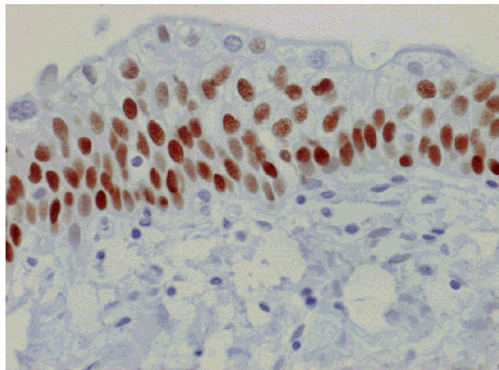
FOXA1



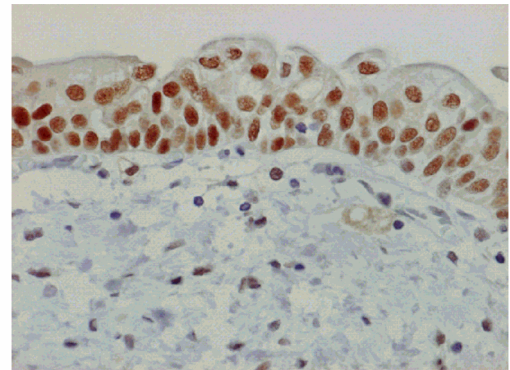
GATA3



GRHL2

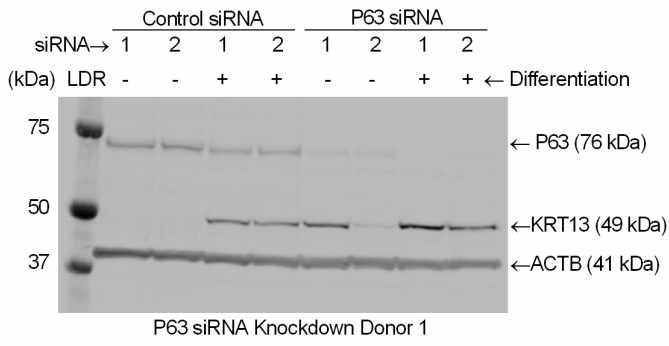
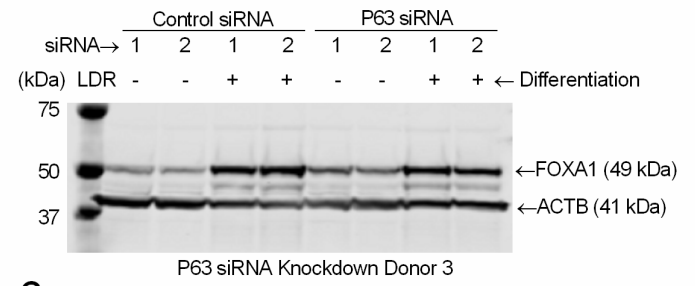
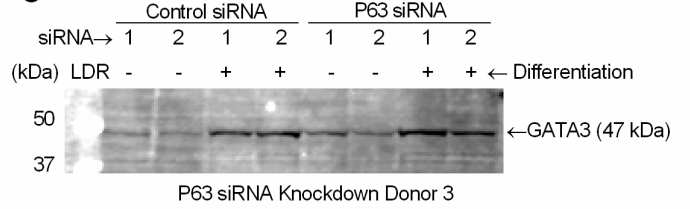
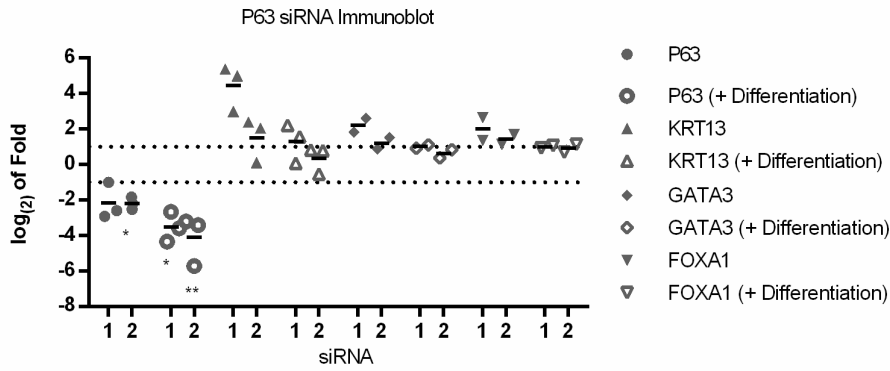
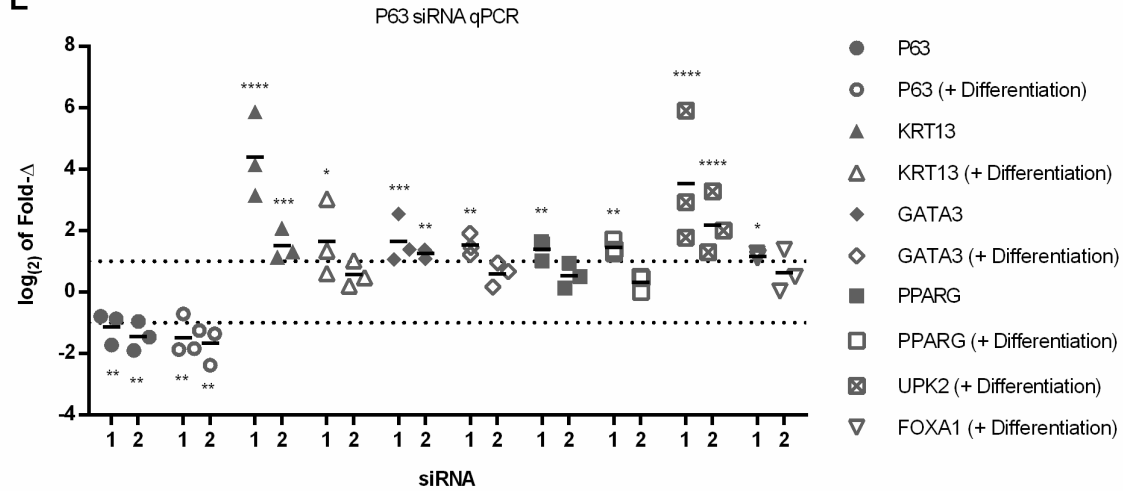
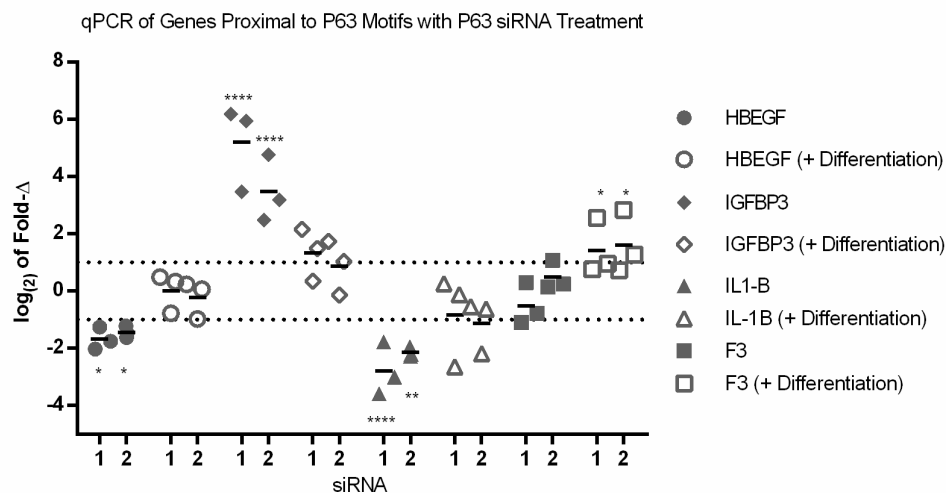


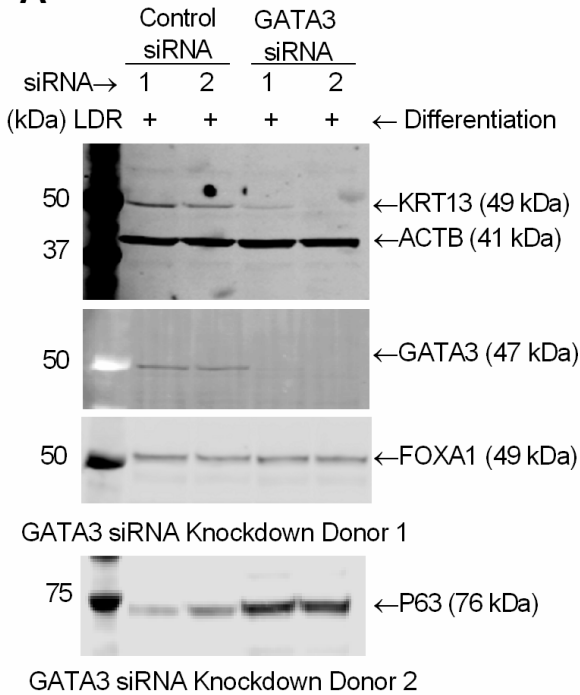
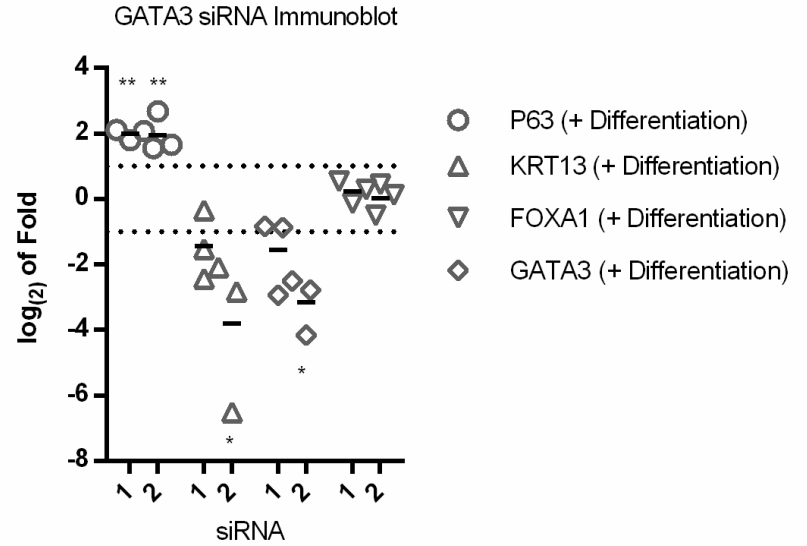
P63



PPARG



**A****B****C****D****E****F**

**A****B****C**