BMC Genomics

Novel long-range regulatory mechanisms controlling *PKD2* gene expression

Stéphanie Moisan[1,2,3*] ⓘ, Stéphanie Levon[1,2], Emilie Cornec-Le Gall[1,3], Yannick Le Meur[4], Marie-Pierre Audrézet[1,3], Josée Dostie[5] and Claude Férec[1,2,3,6*]

Abstract

Background: *Cis*-regulatory elements control gene expression over large distances through the formation of chromatin loops, which allow contact between enhancers and gene promoters. Alterations in *cis*-acting regulatory systems could be linked to human genetic diseases. Here, we analyse the spatial organization of a large region spanning the polycystic kidney disease 2 (*PKD2*) gene, one of the genes responsible of autosomal dominant polycystic kidney disease (ADPKD).

Results: By using chromosome conformation capture carbon copy (5C) technology in primary human renal cyst epithelial cells, we identify novel contacts of the *PKD2* promoter with chromatin regions, which display characteristics of regulatory elements. In parallel, by using functional analysis with a reporter assay, we demonstrate that three DNAse I hypersensitive sites regions are involved in the regulation of *PKD2* gene expression.

Conclusions: Finally, through alignment of CCCTC-binding factor (CTCF) sites, we suggest that these novel enhancer elements are brought to the *PKD2* promoter by chromatin looping via the recruitment of CTCF.

Keywords: ADPKD, *PKD2*, Chromatin organization, Enhancer, Gene expression, Transcription regulation

Background

Autosomal dominant polycystic kidney disease (ADPKD) is one of the most common Mendelian genetic disorders with an incidence of 1/400 to 1/2000 affected individuals worldwide [1]. The disorder is heterogeneous with two main genes, polycystic kidney disease 1 (*PKD1*) and 2 (*PKD2*), which are responsible for most disease cases [2, 3]. While *PKD1* accounts for 75–80% of patients, 10–15% of patients have defects in *PKD2*. A third gene, *GANAB* has recently been identified as accounting for less than 1% of affected patients [4]. Although more than 1200 *PKD1* and almost 200 *PKD2* pathogenic mutations have been reported (ADPKD Mutation Database, http://pkdb.mayo.edu/), a significant proportion of patients with ADPKD symptoms (5–10%) do not appear to carry these pathogenic gene mutations [5–7]. There is significant clinical heterogeneity in the severity of the disease among patients as well as within the same family.

PKD1 and *PKD2* encode two transmembrane glycoproteins, namely polycystin-1 (PC1) and –2 (PC2) which interact via their C-terminal tails [8]. The regulated expression of the PC1/PC2 complex is crucial to understand cyst formation and eventually modulate cyst expression.

The correct expression of a gene relies on the regulatory mechanisms that control it. A part of the regulatory network consists of *cis*-regulatory sequences, such as enhancers and insulators, which may be located far away from their target gene [9–12]. These regulatory elements control gene expression over long distances via the formation of chromatin loops allowing physical proximity between enhancers and gene promoters [13]. Remote regulatory sequences can be identified by mapping DNAse I hypersensitive sites (DHS), which correspond to regions where DNA is more accessible [14–16]. In this context, variations in chromatin organization or alterations of long-range acting elements may impact gene expression. Mutations of these *cis*-acting regulatory sequences can result in human genetic diseases [17, 18], such as aniridia [19] or blepharophimosis syndrome [20, 21]. Therefore, to fully decipher gene expression and genetic

* Correspondence: moisan.stephanie@hotmail.fr; claude.ferec@univ-brest.fr
[1]Institut National de la Santé et de la Recherche Médicale (INSERM), U1078, Brest, Bretagne, France
Full list of author information is available at the end of the article

Moisan *et al. BMC Genomics* (2018) 19:515

Page 2 of 13

disorders, it is particularly important to investigate genetic regulatory networks.

Since the regulation of polycystin is crucial to understand cystogenesis, we decided to explore the long distance *cis*-regulation of *PKD2*. The *PKD2* gene is expressed in several tissues and more particularly in the kidney. In addition to its tissue-specific expression, *PKD2* is also temporally regulated [22]. Moreover, it has been shown that the PC levels can be strictly controlled [23–26]. The mechanisms controlling the tight regulation of *PKD2* expression remain poorly characterized and understood. The *PKD2* promoter does not feature basic regulatory elements such as a TATA or CCAAT box, is GC-rich, and contains multiple putative Sp1, NF-1 and AP-2 protein binding sites [27, 28]. Computational analyses and experimental assays have also described other putative transcription factor binding sites in the *PKD2* promoter region [29–31]. Moreover, several studies have shown that *PKD2* expression is regulated by numerous protein interactions including PKD1 or TRPC1 [32–38]. This control can also occur by other binding protein partners through 5′ and 3′ untranslated regions of *PKD2* mRNA [39–42].

Recent progress in molecular biology and bioinformatics analyses has allowed the development of new techniques to analyse the spatial organization of chromatin. The chromosome conformation capture (3C) technique developed by Dekker and colleagues in 2002 [43] measures the level of contacts between any two defined chromatin regions, and subsequent high-throughput approaches are used to map chromosomal interactions at the scale of the genome. Among these, the chromosome conformation capture carbon copy (5C) measures interaction frequencies within large chromosomal regions at high resolution [44]. We recently used this approach to identify several *CFTR* enhancers [45] and applied it here to a large region spanning the *PKD2* locus in primary human cyst renal epithelial cells. We choose to firstly conduct our long-range *cis*-regulation study with the *PKD2* gene because of challenging molecular analyses of the *PKD1* gene [46]. Approximately 70% of the *PKD1* gene (exons 1–33) is duplicated six times within six pseudogenes (*PKD1P1* to *PKD1P6*), which share a 97.7% sequence identity and a high GC content [47]. We identify novel contacts of the *PKD2* promoter with chromatin regions, which display characteristics of regulatory elements. In parallel, by using functional analysis with a reporter assay we demonstrate that three DHS regions are involved in the regulation of *PKD2* gene expression. Finally, through alignment of CCCTC-binding factor (CTCF) sites, we suggest that these novel enhancer elements are brought to the *PKD2* promoter by chromatin looping via CTCF.
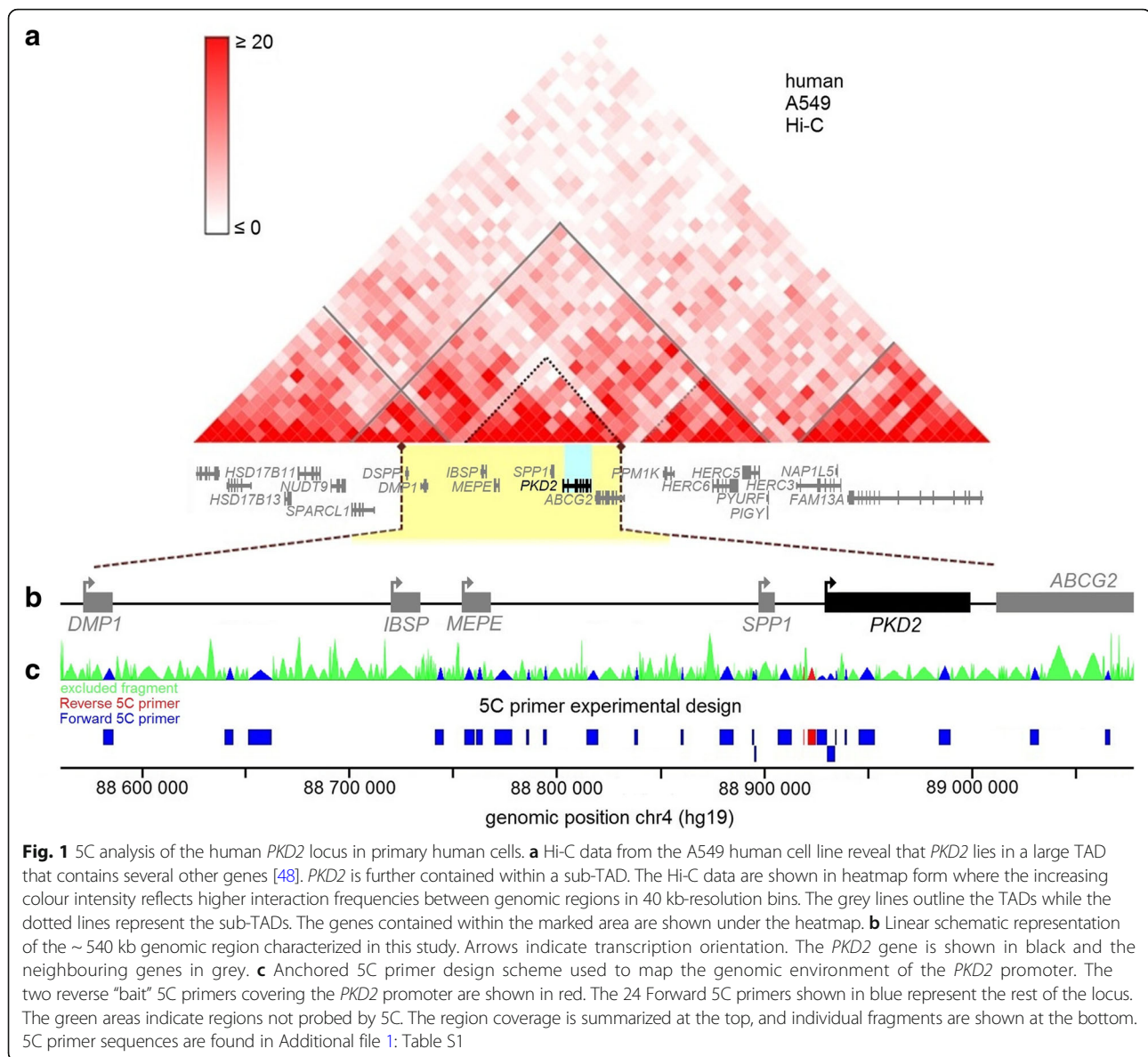
Results

Analysing the human *PKD2* locus with the 5C technology

Very little is known about the regulatory elements that control *PKD2* transcription. To date, no study has examined the involvement of long-range regulatory elements in the mechanisms controlling *PKD2* expression. To pinpoint new regulatory elements involved in the control of *PKD2* expression, we decided to carry out 5C analysis in primary human cells. We used renal cyst epithelial cells (CRC) isolated from patients with ADPKD (carrying a *PKD1* mutation) from the Genkyst cohort who had undergone nephrectomy.

We chose to seek the *PKD2* promoter-interacting partners in a larger locus. To delineate the domain where the *PKD2* promoter interactions are more likely to occur, we used Hi-C data generated from a human lung carcinoma cell line A549 (one of cell lines with highest expression of *PKD2*) [48–50]. Hi-C is a molecular technique, which allows genome-wide quantification of chromosomal interactions in cell populations [51]. This method revealed that chromosomes are organized into topologically associating domains (TADs) corresponding to regions enriched in chromatin contacts [48, 52].

TADs represent folded DNA regions, ranging in size from several hundred kilobases to a few megabases, which is rather conserved across cell types and species [53–55]. TADs are delimited by boundaries, and loci located in different TADs typically interact less frequently than those within the same TAD, suggesting that boundaries act as physical insulators.

TAD boundaries are enriched in CTCF sites [49], and analysis of the A549 Hi-C data revealed that *PKD2* lies within a TAD spreading from the *NUDT9* to the *PYURF* gene (Fig. 1a). *PKD2* itself localizes within a TAD substructure (sub-TAD) [56]. A second TAD overlaps with the domain containing *PKD2* and extends up to the *HSD17B11* gene. Because TADs are largely conserved between cell types, these data suggest that the chromosomal region between these genes may potentially interact with the *PKD2* promoter in different cell lines [48]. We therefore decided to focus our conformational analysis on an ~ 540 kb domain (hg19, chr4: 88,560,000-89,230,000) encompassing *PKD2* and its flanking genes: *DMP1, IBSP, MEPE, SPP1* and *ABCG2* (Fig. 1b). We used 5C-seq to detect and measure chromosomal regions interacting with the *PKD2* promoter in *cis* [44, 57]. The 5C technique has previously been used to detect chromatin organization and map physical contacts between promoters and regulatory elements [45, 56, 58, 59]. Similar to Hi-C, the 5C method quantifies interaction frequencies between DNA regions in fixed cells. The 5C protocol begins with the production of a 3C library where cells are fixed with formaldehyde to capture protein-protein and DNA-protein interactions. The fixed chromatin is then digested with a restriction enzyme

Moisan *et al. BMC Genomics* (2018) 19:515

Page 3 of 13

Fig. 1 5C analysis of the human *PKD2* locus in primary human cells. **a** Hi-C data from the A549 human cell line reveal that *PKD2* lies in a large TAD that contains several other genes [48]. *PKD2* is further contained within a sub-TAD. The Hi-C data are shown in heatmap form where the increasing colour intensity reflects higher interaction frequencies between genomic regions in 40 kb-resolution bins. The grey lines outline the TADs while the dotted lines represent the sub-TADs. The genes contained within the marked area are shown under the heatmap. **b** Linear schematic representation of the ~ 540 kb genomic region characterized in this study. Arrows indicate transcription orientation. The *PKD2* gene is shown in black and the neighbouring genes in grey. **c** Anchored 5C primer design scheme used to map the genomic environment of the *PKD2* promoter. The two reverse "bait" 5C primers covering the *PKD2* promoter are shown in red. The 24 Forward 5C primers shown in blue represent the rest of the locus. The green areas indicate regions not probed by 5C. The region coverage is summarized at the top, and individual fragments are shown at the bottom. 5C primer sequences are found in Additional file 1: Table S1
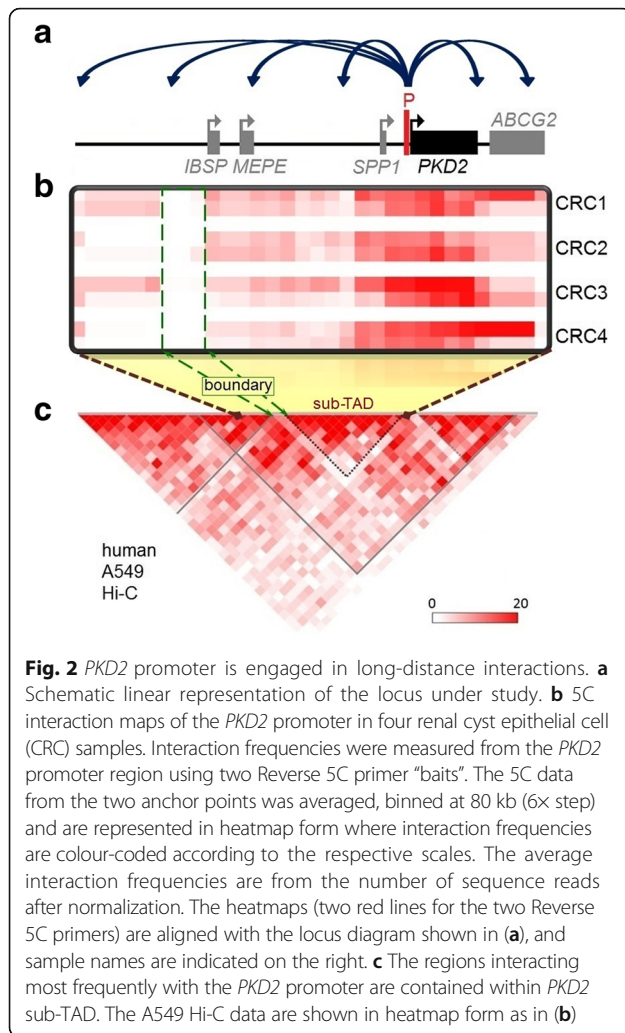
followed by a ligation step that allows the production of pair-wise ligation products at frequencies corresponding to physical proximity between chromatin segments in vivo. The resulting 3C library is converted into a 5C library by a modified ligation-mediated amplification (LMA) procedure where Forward and Reverse 5C primers are annealed and ligated at 3C junctions in a multiplexed setting. 5C library is then amplified by PCR and processed for deep sequencing. We designed 5C primers within our study locus to measure interaction frequencies between the *PKD2* promoter (represented by 2 Reverse 5C primers or "baits") and the rest of the ~ 540 kb domain (Fig. 1c; Additional file 1: Table S1). The 2 reverse primers selected as anchors in this work are corresponding to the most relevant EcoRI restriction fragments upstream to the *PKD2* TSS. Primers of restriction fragments closest to the TSS have been excluded

because of quality parameters, interference and weak results. This 5C design therefore provides information about the genomic environment around the *PKD2* promoter in vivo.

PKD2 promoter is engaged in long-distance interactions

We mapped the interaction profile of the *PKD2* promoter within the ~ 540 kb domain with 5C-seq in four CRC samples isolated from different individuals (Fig. 2a, b). We first observed that the chromatin environment of the *PKD2* promoter captured from either bait is generally similar in all samples (Fig. 2b). Variability is not unexpected in these types of analyses. Chromatin organization is notorious for its plasticity even within cell line samples. This variability has many sources including the different cell cycle phases of a culture sample, and whether or not the cells are

Moisan *et al. BMC Genomics* (2018) 19:515

Page 4 of 13

Fig. 2 *PKD2* promoter is engaged in long-distance interactions. **a** Schematic linear representation of the locus under study. **b** 5C interaction maps of the *PKD2* promoter in four renal cyst epithelial cell (CRC) samples. Interaction frequencies were measured from the *PKD2* promoter region using two Reverse 5C primer "baits". The 5C data from the two anchor points was averaged, binned at 80 kb (6× step) and are represented in heatmap form where interaction frequencies are colour-coded according to the respective scales. The average interaction frequencies are from the number of sequence reads after normalization. The heatmaps (two red lines for the two Reverse 5C primers) are aligned with the locus diagram shown in (**a**), and sample names are indicated on the right. **c** The regions interacting most frequently with the *PKD2* promoter are contained within *PKD2* sub-TAD. The A549 Hi-C data are shown in heatmap form as in (**b**)

exponentially growing and have a high transcription activity, and may come from genetic variations amongst the donors [60]. Interactions were preferentially found within the *PKD2* sub-TAD (Fig. 2b, c; dotted triangle). The highest levels of contacts were observed mostly within the *PKD2* gene and with the upstream and downstream regions containing *SPP1* and *ABCG2* genes, respectively. Importantly, we also detected long-range contacts between the *PKD2* promoter and multiple distal chromatin regions, particularly with one located 5′ of the *IBSP* and *MEPE* genes (Fig. 2b). As expected, we observed a reduction in the interaction frequency (highlighted in dashed green box), which corresponds to a TAD boundary (Fig. 2b, c). In fact, topological domains are separated by tight regions, *boundary regions*, where chromatin interactions are less frequent [48]. Altogether, these results suggest that the *PKD2* promoter engages in long-range contacts with several putative regulatory elements, which could be important in regulating its expression.

PKD2 contact regions display characteristics of regulatory elements

Some of the 5′ and 3′ regions engaged in long-range contacts with the *PKD2* promoter might correspond to DNA elements important for its regulation. We therefore tested if some of these regions displayed features of enhancers by using available databases for different markers of regulatory elements. We aligned our 5C interaction profiles from the four CRC samples with datasets from the human Renal Proximal Tubule Epithelial Cells (RPTECs). We assumed that the regulatory elements controlling *PKD2* in this cell line should be, at least in part, similar to the regulatory network in cyst epithelial cells. We specifically looked for the presence of DHS or H3K27Ac marks because these often reflect the presence of regulatory elements.

We found four strong chromatin interactions between *PKD2* promoter and the rest of the locus of interest in the four CRC samples (Fig. 3a, "I to IV"). These regions coincide with DHS and/or H3K27Ac marks (acetylation of the 27th lysine residue on histone protein 3) in RPTECs (Fig. 3b, c). Three contacts involve sequences within the 5′ region (Fig. 3a, 'I to III') and one is within the *PKD2* gene body (Fig. 3a, 'IV'). These interacting regions were located between − 151 kb (for region I) and + 5 kb (for region IV) from the *PKD2* transcriptional start site (TSS).

Several DHS regions show enhancer activities

A common attribute of functional elements is their hypersensitivity to DNAse I, which corresponds to nucleosome-free regions where regulatory factors are bound. To further assess whether the region selected around *PKD2* for this study contains bona fide regulatory elements, we aligned available DHS data from A549 cells along our region and chose to test eight DHS fragments for enhancer activities against the *PKD2* promoter in a reporter assay (Fig. 4a, 'A to H'). For this analysis, we first prepared a 'P$_{PKD2}$' construct by subcloning a 1169 bp *PKD2* promoter fragment into the pGL3-Basic vector upstream of a modified firefly luciferase coding region optimized for analysing transcriptional activity in eukaryotic cells. Regions A to H were then PCR amplified and inserted into P$_{PKD2}$ (Additional file 2: Table S2). These different constructs were individually co-transfected into A549 cells with a beta-galactosidase plasmid as a control for transfection efficiency. Firefly luciferase expression was measured after 48 h and was normalized against the *PKD2* promoter construct alone (P$_{PKD2}$), which was set to 1.

As shown in Fig. 4b, three of our DHS fragments significantly enhanced *PKD2* promoter activity. The fragment encompassing DHS H (hg19, chr4:89,099,146-89,100,467), located at + 170 kb from *PKD2* TSS, displayed a strong
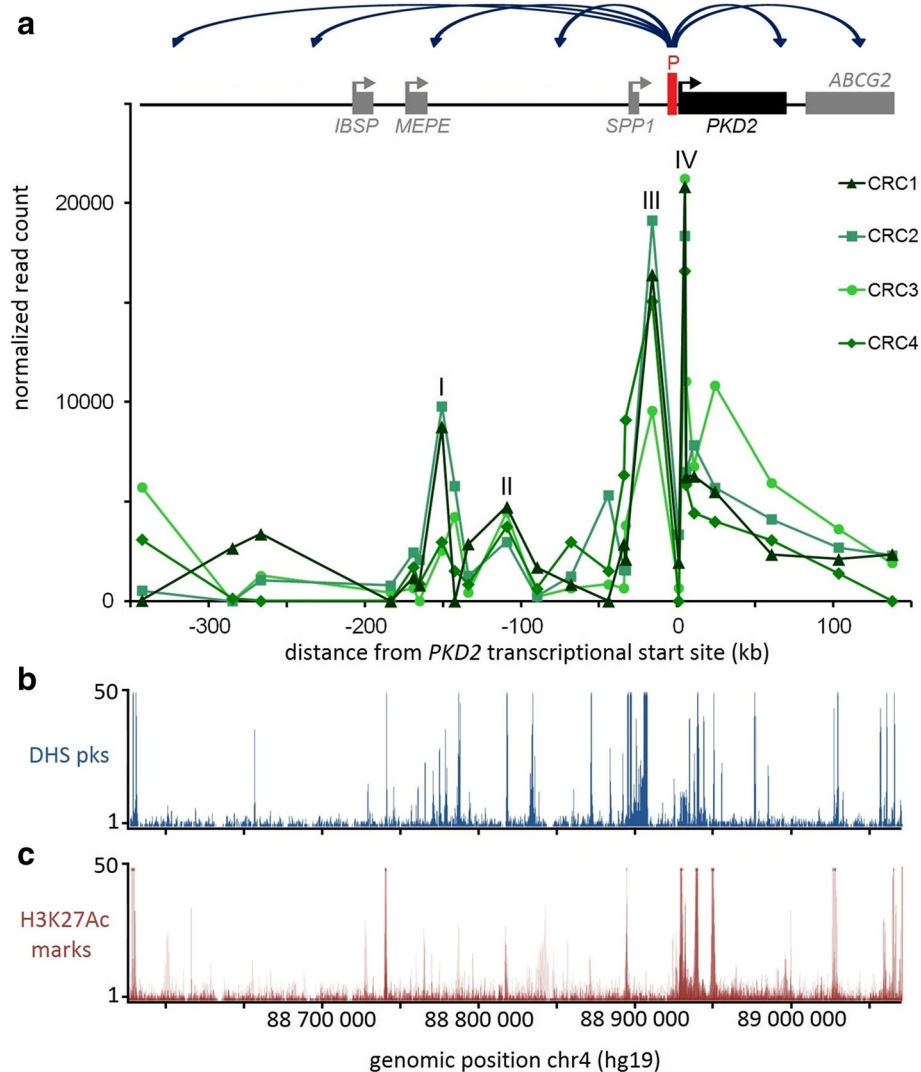
Moisan *et al. BMC Genomics* (2018) 19:515

Page 5 of 13

Fig. 3 *PKD2* interacting regions display characteristics of regulatory elements. **a** 5C chromatin interaction profiles of the *PKD2* promoter region in four CRC samples. Interaction frequency (*y-axis*) is correlated with the position from the transcriptional start site of *PKD2* (*x-axis*). The four strong contact regions are numerated I to IV. **b** Alignment of DNAse I hypersensitivity data from RPTEC over the region characterized. **c** Alignment of H3K27Ac mark data from RPTEC over the region characterized

enhancer function with more than ten-fold effect on the *PKD2* promoter. The DHS region F (hg19, chr4: 88,938,944-88,940,420), located at + 10 kb from *PKD2* TSS, increased the promoter activity by 5.1-fold and DHS C (hg19, chr4:88,895,792-88,897,206), located at − 32 kb from *PKD2* TSS, had a more modest enhancer activity with an almost three-fold effect. However, we observed no evidence of cooperation among these three elements (data not shown). The other DHS regions did not show any enhancer activity in A549 cells.

Formation of a *PKD2* "active chromatin hub"

The four regions (Fig. 5a, 'I to IV') that we found to strongly interact with the *PKD2* promoter in four CRC samples with 5C analysis, correspond to DHS fragments

A, B, D and E, respectively, and do not show *PKD2* enhancer activity. However, some of these regions are localized at more or less close proximity to strong CTCF peaks in RPTECs (Fig. 5b). CTCF functions as a genome architectural protein organizer that contributes to create local chromatin hubs, to join clusters of genes with coordinated expression and to promote communication between regulatory elements and their corresponding promoters [61]. A Subset of chromatin loops that are formed by a CTCF-CTCF homodimer brings sequences that are far apart in the linear genome into close proximity [62]. Thus, CTCF-associated interactions potentially allow extensive cross-talk between promoters and distant enhancers [63]. CTCF is sometimes referred to as an enhancer facilitator [61]. CTCF binding is largely
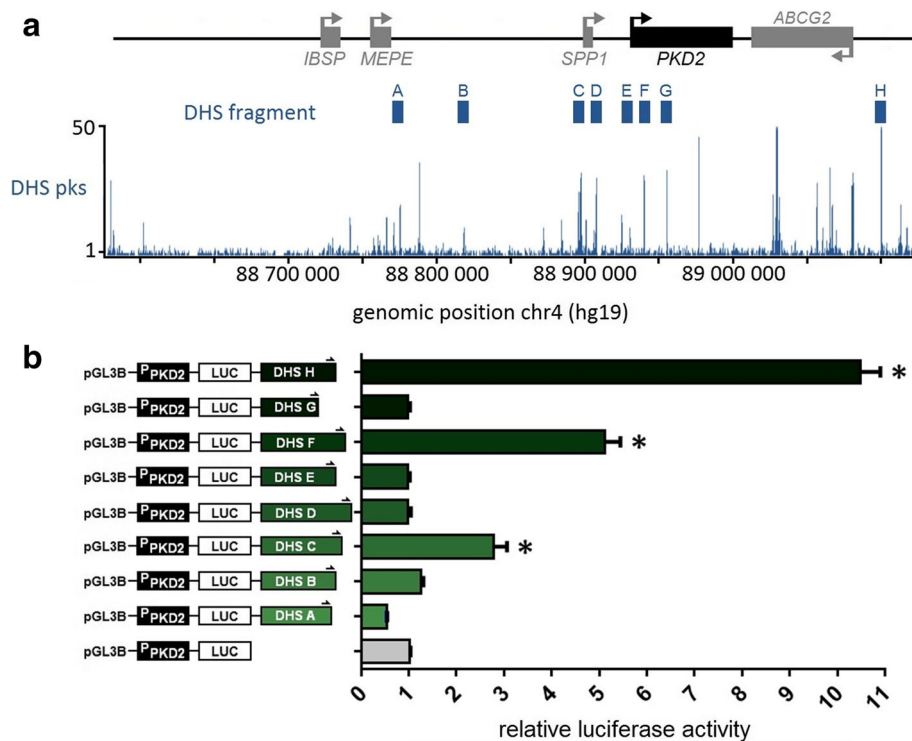
Moisan *et al. BMC Genomics* (2018) 19:515

Page 6 of 13

Fig. 4 Several DHS regions show enhancer activities. **a** Alignment of DNAse I hypersensitivity data from A549 cells over the schematic linear representation of the region. Gene orientation is indicated with arrows. The cloned DHS elements are indicated with blue boxes at the top and identify open chromatin sites. **b** A549 cells were transfected with pGL3B luciferase reporter constructs containing the *PKD2* basal promoter (P$_{PKD2}$; 1169 bp) and fragments of DHS elements of the region A to H. Luciferase data are shown relative to the *PKD2* basal promoter vector (=1). Error bars represent SEM ($n = 9$), * $P < 0.0001$ using unpaired t-tests

invariant across cell types in mammals [64], and so these nucleosome-free regions could be implicated in the formation of a *PKD2* 'active chromatin hub' due to the recruitment of CTCF factors. Moreover, enhancers C and F are located directly near DHS fragments D and E, which are respectively not so far and at close proximity of CTCF binding sites. Thus, we suggest that these enhancers activate the *PKD2* promoter by achieving physical proximity via CTCF-CTCF binding and looping of the intervening chromatin.

Discussion

The crucial step in the development of cysts in ADPKD is linked to the amount of functional PC1 or PC2, and a reduction in the amount of these functional polycystins (also called the dosage model) leads to cystogenesis [65]. It is possible that the threshold could vary between tissues and cell types as well as with the type of mutated allele. This highlights the importance of understanding the regulation and expression of polycystins in order to discover new targets to increase their expression and possibly decrease cyst formation and slow disease progression.

Because the regulatory mechanisms of *PKD2* transcription remain largely uncharacterized and some cases of ADPKD have no genotypic explanation [7], it is essential to investigate alternative control mechanisms, particularly long-range regulation. We used 5C technology to identify regions that interact specifically with the *PKD2* promoter and may be involved in regulating its expression. We decided to study these interactions in primary cells to avoid altered expression patterns of cell lines. For this purpose, we used human renal epithelial tissues obtained from nephrectomy.

Despite some minor variations among cell samples, our results show that the *PKD2* promoter is engaged in numerous long-range interactions (Fig. 2a). These chromatin contacts are mostly located within the *PKD2* sub-TAD (Fig. 2b, c). These interacting partners lie downstream and upstream of extragenic regions and within intronic regions, preferentially within the *PKD2* gene as well as just upstream and downstream of regions containing *SPP1* and *ABCG2* genes. Interactions of the *PKD2* promoter with neighbouring regions decrease substantially with distance, especially at a region corresponding to a TAD boundary (dashed green box, Fig. 2b, c). This is consistent with studies that recently showed that

Moisan *et al. BMC Genomics* (2018) 19:515

Page 7 of 13

Fig. 5 Formation of a *PKD2* "active chromatin hub". **a** Alignment of 5C chromatin interaction profiles of the *PKD2* promoter region in four CRC samples with the cloned DHS elements to highlight the position of enhancers C, F and H relative to interacting regions. **b** Alignment of CTCF binding sites from RPTEC to highlight the position of enhancers and interacting regions relative to CTCF peaks

chromatin interactions within *boundary regions* is practically non-existent [48, 52].

Our 5C data revealed four strong interacting regions that correlate with marks of regulatory elements (Fig. 3a-c). For example, we detected contacts between the *PKD2* promoter and a region located ~ 150 kb upstream (Fig. 3a, region I) containing DHS sites (Fig. 3b) and a high CTCF binding site (Fig. 5a, b).

We also found interactions between the promoter and an element located at ~ 110 kb from the *PKD2* TSS in 5′ region (region II), which coincides with DHS and H3K27Ac marks. Furthermore, we detected interactions with two others regions located at 16 kb upstream and 5 kb downstream of the *PKD2* TSS (region III and IV), which correspond to strong DHS, H3K27Ac marks and are at proximity of CTCF binding sites.

In addition, by using reporter assays where luciferase expression is driven by a *PKD2* promoter fragment, we found that three DHS regions act as enhancers (Fig. 4b). The region H, located at + 170 kb of the *PKD2* promoter show the strongest enhancer activity with more than 10-fold increase of the *PKD2* expression in A549 cells. DHS C at approximately − 32 kb and DHS F at approximately + 10 kb from the *PKD2* TSS have more modest enhancer activities, increasing the promoter activity by 2.8- and 5.1-fold, respectively. These three regions span sequences found to bind different types of transcription factors and chromatin remodelers in an ENCODE compilation of 161 factors with Factorbook motifs in 91 cell types (Additional file 3: Figure S1). To identify the transcription factors driving enhancer activities, these DHS regions were evaluated in subfragments. These segments

Moisan *et al. BMC Genomics* (2018) 19:515

Page 8 of 13

showed a more modest activity even no enhancement compared with entire regions (data not shown). These results suggest that combinatorial regulatory elements would be involved [66, 67]. For example, transcription factors FOS and JUND could act cooperatively in region F.

In addition, others regions (Fig. 3, regions I to IV that correspond to DHS A, B, D and E, respectively) that correlate with regulatory markers and interact with the *PKD2* promoter without increasing its expression, may still function in the regulation of *PKD2* through other mechanisms or by acting in a combinatorial fashion. In particular, enhancers C and F are located directly near DHS fragments D and E which are respectively not so far and at close proximity of CTCF binding sites, suggesting that the increase of the *PKD2* promoter by these *cis*-acting regulatory elements could depend on the active chromatin looping driven by these insulator factors which bring them into spatial proximity.

In conclusion, we identified a new set of *PKD2* distal *cis*-regulatory elements in primary human renal epithelial cells. We proposed a looping model in the organization of the *PKD2* locus in renal cells via the formation of a chromatin loop with CTCF-CTCF binding around the gene.

It is evident that other sequence variabilities in the coding region, in the promoter sequence or at a distance from the promoter region of the *PKD* genes are important to explain the regulation of the promoter activity, which can influence cyst formation [68].

Future work will be aimed at identifying this long-range regulation of *PKD2* expression and, in particular, the trans-acting factors that mediate it in greater detail.

Identification of long-range acting regulatory elements that affect expression of genes may also be important for genetic diagnosis of ADPKD. It will also be very interesting to study *cis*-regulation with the *PKD1* promoter and potential co-regulation between these two genes. Mutations in these regulatory elements can also explain many cases of patients who have either a hypomorphic allele in the *PKD1* or *PKD2* gene coding sequences or none at all. Moreover, genetic alterations within remote-acting regions could lead to *PKD* dysregulation, which might explain variability in phenotypic features of ADPKD, such as the development of renal insufficiency [69]. Some regulatory variants have already been associated to extensive phenotypic variability. Indeed, recently non-coding variants have been related with cystic fibrosis disease traits [70]. Thus, these *PKD2* cis-regulatory elements could be sequenced in ADPKD patients in whom no mutation was found in the *PKD* gene coding sequences or patients with extreme clinical phenotypes. Maybe, with time, these regions would be analysed by whole-genome sequencing [71].

Conclusion

In conclusion, we reported for the first time the use of 5C-seq technology in primary human cells to analyse the chromatin conformation of a locus involved in ADPKD and identified new regulatory elements. We demonstrated that three regions surrounding the *PKD2* gene interact with the promoter and activate its expression in renal epithelial cells. We proposed a regulatory looping model in the 3D-organization of the *PKD2* locus in renal cells via the formation of a chromatin loop thanks to the recruitment of CTCF.

These findings provide valuable information on the basic insights of the regulation of this gene, and the description of *PKD2* regulatory elements may be important for gene targeting constructs for therapy development. They also afford new targets for mutation screening in genetic diagnosis of ADPKD.

Methods
Cell collection and culture
Human renal tissues were obtained from PKD1 patients with ADPKD (carrying no *PKD2* mutation) from the Genkyst cohort who had undergone nephrectomy. Several cyst tops were individually dissected with a scalpel. Each cyst was rinsed thoroughly in ice-cold phosphate buffered saline (PBS) and then placed in a separate tube containing 0.15% Pronase E. Cysts were incubated for 24 h at 4 °C to allow separation of cells from the cyst wall. Dissociation was stopped by addition of foetal bovine serum (FBS). The cyst walls stripped of cells were removed by filtration, and the cells were washed twice in PBS. Renal cyst epithelial cell (CRC) suspensions were centrifuged for 5 min at 2200 *g* at 4 °C. Cell pellets were resuspended in a selected and optimized cell culture media for human renal epithelial cells (REGM; Lonza), and cells were cultured following the manufacturer's instructions.

A549 human lung carcinoma cells [72] were grown in Dulbecco's modified eagle medium (DMEM; Lonza) with 10% FBS.

All cells were grown on plastic at liquid interface at 37 °C in 5% CO_2 saturated humid air.

Cell isolation and fixation for 5C analysis
CRC cells were fixed with 1.5% formaldehyde for 10 min at room temperature. Crosslinking was stopped with glycine (125 mM final concentration) by incubating 5 min at room temperature followed with 15 min on ice. Cells were scraped and centrifuged at 400 *g* for 10 min at 4 °C. Supernatants were removed and the cell pellets were quick-frozen on dry ice.

3C library preparation
The fixed CRC cell pellets were treated as previously described [73, 74]. Briefly, 10 million cells were incubated

Moisan *et al. BMC Genomics* (2018) 19:515

Page 9 of 13

in lysis buffer (10 mM Tris (pH 8.0), 10 mM NaCl, 0.2% NP-40, supplemented with fresh protease inhibitor cocktail) for 10 min on ice. Cells were then disrupted on ice with a Dounce homogenizer (pestle B; 2×20 strokes). Nuclei were recovered by centrifugation, washed twice with 1X EcoRI buffer (NEB) and resuspended in 100 μl of 1X EcoRI buffer.1X EcoRI buffer (337 μl) was added to 50 μl of cell suspension, and the mixture was incubated for 10 min at 65 °C with 0.1% SDS final concentration (38 μl). Triton X-100 (44 μl of 10% Triton X-100) was added before overnight digestion with EcoRI (400 U). The restriction enzyme was inactivated on the next day by adding 86 μl of 10% SDS and incubating 30 min at 65 °C. Samples were then individually diluted into 7.62 ml of ligation mix (750 μl of 10% Triton X-100, 750 μl of 10X ligation buffer, 80 μl of 10 mg/ml BSA, 80 μl of 100 mM ATP and 3000 Cohesive end Units of T4 DNA ligase). Ligation was carried out at 16 °C for 2 h.

The 3C libraries were next incubated overnight with 45 μL of Proteinase K (10 mg/ml) at 55 °C.The DNA was purified by phenol-chloroform extraction and precipitated with 3 M NaOAc pH 5.2 (800 μl) and cold ethanol. After at least 1 h at − 80 °C, the DNA was recovered by centrifugation, and the pellets were washed with cold 70% ethanol and then resuspended in 400 μl of 1X TE pH 8.0. A second phenol-chloroform extraction and precipitation with 3 M NaOAc pH 5.2 (40 μl) and cold ethanol was performed. DNA was recovered by centrifugation and washed eight times with cold 70% ethanol. The pellets were finally dissolved in 100 μl of 1X TE pH 8.0 and incubated with RNAse A (1 μl at 10 mg/ml) for 15 min at 37 °C.

5C primer and library design
5C primers covering the *PKD2* region (hg19, chr4: 88,560,000-89,230,000), the *ERCC3* region (hg19, chr2: 128,014,866-128,051,752) and the *ENr313* region (hg19, chr16:62,276,449-62,776,448) were designed using "my5C. primer" [75] with the following parameters: optimal primer length of 30 nt, optimal TM of 60 °C, default primer quality parameters (mer: 800, U-blast: 3, S-blasr: 50). Primers were not designed for large (> 20 kb) restriction fragments. Low complexity and repetitive sequences were excluded from our experimental design such that not all fragments could be probed in our assays. Primers with several genomic targets were also removed.

The universal A-key (CCATCTCATCCCTGCGTGT CTCCGACTCAG-(5C-specific)) and the P1-key tails ((5C-specific)-ATCACCGACTGCCCATAGAGAGG) were added to the Forward and Reverse 5C primers, respectively. Reverse 5C primers were phosphorylated at their 5′ ends. An anchored 5C design was used for the *PKD2* locus, 2

Reverse 5C primers targeted the *PKD2* promoter while 24 Forward 5C primers covered the surrounding region.

For the control regions, an alternating 5C design was used: alternating Forward and Reverse 5C primers covering *ERCC3* and *ENr313* regions were used to generate the 5C libraries. This design used 31 primers (2 Forward / 3 Reverse for the *ERCC3* region, 13 Forward / 13 Reverse *ENr313* region). All 5C primer sequences are listed in Additional file 1: Table S1.

5C library preparation
5C libraries were prepared and amplified with the A-key and P1-key primers following a procedure described previously [57]. Briefly, 3C libraries were first titrated by PCR for quality control (single band, absence of primer dimers, etc.) and to verify that contacts are amplified at frequencies similar to what is usually obtained from comparable libraries (same DNA amount from the same species and karyotype) [73, 76, 77].

The 5C primer stocks (20 μM) were diluted individually in water on ice, and mixed to a final concentration of 0.002 μM. The mixed primers were combined with annealing buffer (10X NEBuffer 4, New England Biolabs Inc.) on ice in reaction tubes. Salmon testis DNA was added to each 5C reaction, followed by the 3C libraries and water for a final volume of 10 μl. Samples were denatured at 95 °C for 5 min, and annealed at 50 °C for 16 h before ligation with Taq DNA ligase (10 U) for one hour. Each ligation reaction was then PCR-amplified individually with primers against the A-key and P1-key primer tails. We used 35 cycles based on dilution series showing linear PCR amplification within that cycle range. The 5C PCR products of corresponding 3C libraries were pooled before purifying the DNA on MinElute columns (Qiagen).

The 5C libraries were quantified on agarose gel and diluted to 0.012 ng/μl (for Ion PGM™ Template OT2 200 Kit). A 25 μL volume of the diluted 5C library was used for sequencing with an Ion PGM™ sequencer. Samples were sequenced onto Ion 314™ Chips v2 or Ion 316™ Chips following the Ion PGM™ Template OT2 200 Kit and Ion PGM™ Sequencing 200 Kit v2 protocols recommended by the manufacturer (Life Technologies™).

5C data analyses
Analysis of the 5C sequencing data were performed as described previously [57, 59]. The sequencing data were processed through a Torrent 5C data transformation pipeline on our local instance of Galaxy as previously described [57, 78] (http://galaxy.bci.mcgill.ca/galaxy/). This analysis generates an excel sheet containing interaction frequency lists (IFL) as well as a text file, which were used to visualize results using "my5C-heatmap" [75].

Moisan *et al. BMC Genomics* (2018) 19:515

Page 10 of 13

Normalization between different libraries was done first by read count and then by using the compaction profiles for the *ERCC3* region and the gene desert region *ENr313* after setting one sample (CRC) as a reference.

The *ERCC3* and *ENr313* 5C data for each sample were normalized by dividing the number of reads of each 5C contact by the total number of reads from the corresponding sequence run. A ratio, calculated by dividing these normalized data by normalized data from a reference sample (CRC), was applied to the corresponding raw data of the study region for each sample.

5C data was averaged, optimally binned at 80 kb (6× step) and are represented in heatmap form where interaction frequencies are colour-coded according to the respective scales. The average interaction frequencies are from the number of sequence reads after normalization. Data was binned based on distance for several reasons. First, by binning the data and re-dividing it into defined genomic distances (or "steps"), 5C signals are averaged along the region such that 5C contacts with greater error rates will not be seen as false positive or negative. This is particularly important given that the data is displayed on a color scale and that the eye does not detect signal linearly. By binning and re-dividing we also avoid displaying gaps which my5C inserts as grey lines that detracts attention from the actual findings. At the same time, this transformation allows us to align the data directly onto the genomic region without having to use "true size" - a display of the data where bin sizes are the same as the fragments sizes and which can enhance or reduce contact frequency to the eye. "True size" displays also insert grey lines in the heatmaps.

Databases and URLs

The Hi-C data from human A549 were downloaded from computational and functional Genomics/Epigenomics Yue Lab website at http://promoter.bx.psu.edu/hi-c/view.php. The Hi-C data shown in a heatmap format were binned at 40 kb. TAD and sub-TADs were outlined manually based on clustering of interaction frequencies.

DHS peaks, CTCF binding sites datasets of H3K27Ac marks for RPTEC and A549 and transcription factor ChIP-seq data of 161 factors from 91 cell types were obtained and visualized using the UCSC genome browser (http://genome.ucsc.edu) [79].

The "my5C-primer" and "my5C-heatmap" bioinformatics tools are found at http://3dg.umassmed.edu/my5Cheatmap/heatmap.php.

Plasmid constructs

All the cloning steps were done using the "In-fusion®" strategy from Clontech. Using the pGL3-Basic Vector (Promega), the 5′-flanking region of the *PKD2* gene (1169 bp, "P_{PKD2}") was cloned upstream of the firefly luciferase cDNA at the Hind III site. The candidate enhancer elements (A to H) were amplified and inserted downstream of the P_{PKD2} construct.

All the inserted fragments were verified by sequencing. The PCR primers used to amplify the *PKD2* promoter and candidate enhancer sequences are shown in Additional file 2: Table S2.

Luciferase assay

Cells (2.5×10^5) were seeded in 6-well plates. Transfections were done 24 h later with the transit 2020 reagent (Mirus). 800 ng of the P_{PKD2} construct and 200 ng of the pCMV-LacZ construct (as internal control) were used for each condition. Every condition was carried out in triplicates.

The cells were washed once with 1X PBS and lysed with Passive lysis buffer (Promega) 48 h post-transfection. Cell lysates were clarified by centrifugation at 15,000 g for 5 min at 4 °C. 20 μL of each protein extract was used to assay the luciferase activity and 25 μL for beta-galactosidase activity. We used Promega reagents and a multiwell plate reader Varioskan (Thermo Fisher). The results were presented as relative luciferase activity with the P_{PKD2} construct activity equal to 1. Significance determination of the increased luciferase activity was performed using unpaired *t*-tests using the GraphPad Prism® software.

Additional files

Additional file 1: Table S1. 5C primer sequences for the *PKD2*, *ERCC3* and ENr313 regions. (PDF 75 kb)

Additional file 2: Table S2. PCR primer sequences used for cloning into the luciferase reporter construct (5′ - 3′). (PDF 12 kb)

Additional file 3: Figure S1. Enhancer regions overlap with regulatory element binding sites. Enhancer regions C (A), F (B) and G (C) overlap with several transcription factor or chromatin remodeler binding sites with Factorbook motifs identified by ChIP-seq in 91 different cell lines. (TIF 385 kb)

Moisan *et al. BMC Genomics* (2018) 19:515

Page 11 of 13

Availability of data and materials

The Hi-C data from human A549 were downloaded from computational and functional Genomics/Epigenomics Yue Lab website. The Hi-C data shown in a heatmap format were binned at 40 kb (http://promoter.bx.psu.edu/hi-c/view.php?species=human&assembly=hg19&source=inside&tissue=A549&type=raw-rep1&resolution=40&c_url=&transfer=&chr=chr4&start=87720000&end=90000000&sessionID=&browser=none).
DHS peaks, CTCF binding sites datasets of H3K27Ac marks for RPTEC and A549 and transcription factor ChIP-seq data of 161 factors from 91 cell types were obtained and visualized using the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr4%3A88576500-89070500&hgsid=676703611_3vY26dSGrdehcr939HtB3zszv3tE) [79].
The "my5C-primer" and "my5C-heatmap" data are available at https://docs.google.com/spreadsheets/d/1w%2D-3mF1cbf6za6uOxWU6GnpM8XM5XN0Axp0rB24cOpM/edit?usp=sharing.

Authors' contributions

SM: Designed and supervised the study, carried out the experiments, analyzed the data, wrote the paper. SL: Carried out 5C experiments, analyzed the data. ECLG: Provided material support, edited the paper. YLM: Provided material support. MPA: Provided material support, edited the paper. JD: Provided data analysis software, edited the paper. CF: Supervised lab work, edited the paper. All authors read and approved the manuscript.

Ethics approval and consent to participate

This study with renal tissues from the Genkyst cohort was conducted with the approval of the Comité de Protection des Personnes Ouest VI (DC 2010 1125), of CHU Brest Cavale Blanche (CHU Cavale Blanche - 29,609 BREST CEDEX – FRANCE). Participants gave informed written consent and received clinical information.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Institut National de la Santé et de la Recherche Médicale (INSERM), U1078, Brest, Bretagne, France. [2]Faculté de Médecine et des Sciences de la Santé, Université de Bretagne Occidentale (UBO), Brest, Bretagne, France. [3]Laboratoire de Génétique Moléculaire et d'Histocompatibilité, Centre Hospitalier Régional Universitaire (CHRU), Hôpital Morvan, Brest, Bretagne, France. [4]Service de néphrologie, Centre Hospitalier Régional Universitaire (CHRU), Brest, Bretagne, France. [5]Department of Biochemistry and Goodman Cancer Research Center, McGill University, Montréal, Québec, Canada. [6]Etablissement Français du sang (EFS), Brest, Bretagne, France.

References

1. Harris PC, Torres VE. Polycystic kidney disease. Annu Rev Med. 2009;60:321–37.
2. The European Polycystic Kidney Disease Consortium. The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. Cell. 1994;77(6):881–94.
3. Mochizuki T, Wu G, Hayashi T, Xenophontos SL, Veldhuisen B, Saris JJ, Reynolds DM, Cai Y, Gabow PA, Pierides A, et al. PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. Science. 1996;272(5266):1339–42.
4. Porath B, Gainullin VG, Cornec-Le Gall E, Dillinger EK, Heyer CM, Hopp K, Edwards ME, Madsen CD, Mauritz SR, Banks CJ, et al. Mutations in GANAB, encoding the glucosidase IIalpha subunit, cause autosomal-dominant polycystic kidney and liver disease. Am J Hum Genet. 2016;98(6):1193–207.
5. Audrezet MP, Cornec-Le Gall E, Chen JM, Redon S, Quere I, Creff J, Benech C, Maestri S, Le Meur Y, Ferec C. Autosomal dominant polycystic kidney disease: comprehensive mutation analysis of PKD1 and PKD2 in 700 unrelated patients. Hum Mutat. 2012;33(8):1239–50.
6. Harris PC, Rossetti S. Molecular diagnostics for autosomal dominant polycystic kidney disease. Nat Rev Nephrol. 2010;6(4):197–206.
7. Rossetti S, Consugar MB, Chapman AB, Torres VE, Guay-Woodford LM, Grantham JJ, Bennett WM, Meyers CM, Walker DL, Bae K, et al. Comprehensive molecular diagnostics in autosomal dominant polycystic kidney disease. J Am Soc Nephrol. 2007;18(7):2143–60.
8. Fedeles SV, Tian X, Gallagher AR, Mitobe M, Nishio S, Lee SH, Cai Y, Geng L, Crews CM, Somlo S. A genetic interaction network of five genes for human polycystic kidney and liver diseases defines polycystin-1 as the central determinant of cyst formation. Nat Genet. 2011;43(7):639–47.
9. Kleinjan DA, van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet. 2005; 76(1):8–32.
10. Dekker J. Gene regulation in the third dimension. Science. 2008;319(5871): 1793–4.
11. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414): 57–74.
12. West AG, Fraser P. Remote control of gene transcription. Hum Mol Genet. 2005;14(1):R101–11.
13. Marsman J, Horsfield JA. Long distance relationships: enhancer-promoter communication and dynamic gene transcription. Biochim Biophys Acta. 2012;1819(11–12):1217–27.
14. Harmston N, Lenhard B. Chromatin and epigenetic features of long-range gene regulation. Nucleic Acids Res. 2013;41(15):7185–99.
15. Zentner GE, Scacheri PC. The chromatin fingerprint of gene enhancer elements. J Biol Chem. 2012;287(37):30888–96.
16. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010;107(50):21931–6.
17. Kleinjan DJ, Coutinho P. Cis-ruption mechanisms: disruption of cis-regulatory control as a cause of human genetic disease. Brief Funct Genomic Proteomic. 2009;8(4):317–32.
18. Crutchley JL, Wang XQ, Ferraiuolo MA, Dostie J. Chromatin conformation signatures: ideal human disease biomarkers? Biomark Med. 2010;4(4):611–29.
19. Kleinjan DA, Seawright A, Schedl A, Quinlan RA, Danes S, van Heyningen V. Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6. Hum Mol Genet. 2001;10(19):2049–59.
20. Beysen D, Raes J, Leroy BP, Lucassen A, Yates JR, Clayton-Smith J, Ilyina H, Brooks SS, Christin-Maitre S, Fellous M, et al. Deletions involving long-range conserved nongenic sequences upstream and downstream of FOXL2 as a novel disease-causing mechanism in blepharophimosis syndrome. Am J Hum Genet. 2005;77(2):205–18.
21. D'Haene B, Attanasio C, Beysen D, Dostie J, Lemire E, Bouchard P, Field M, Jones K, Lorenz B, Menten B, et al. Disease-causing 7.4 kb cis-regulatory deletion disrupting conserved non-coding sequences and their interaction with the FOXL2 promotor: implications for mutation screening. PLoS Genet. 2009;5(6):e1000522.
22. Markowitz GS, Cai Y, Li L, Wu G, Ward LC, Somlo S, D'Agati VD. Polycystin-2 expression is developmentally regulated. Am J Phys. 1999;277(1 Pt 2):F17–25.
23. Park EY, Sung YH, Yang MH, Noh JY, Park SY, Lee TY, Yook YJ, Yoo KH, Roh KJ, Kim I, et al. Cyst formation in kidney via B-Raf signaling in the PKD2 transgenic mice. J Biol Chem. 2009;284(11):7214–22.
24. Wu G, D'Agati V, Cai Y, Markowitz G, Park JH, Reynolds DM, Maeda Y, Le TC, Hou H Jr, Kucherlapati R, et al. Somatic inactivation of Pkd2 results in polycystic kidney disease. Cell. 1998;93(2):177–88.
25. Lu W, Peissel B, Babakhanlou H, Pavlova A, Geng L, Fan X, Larson C, Brent G, Zhou J. Perinatal lethality with kidney and pancreas defects in mice with a targetted Pkd1 mutation. Nat Genet. 1997;17(2):179–81.
26. Igarashi P, Somlo S. Genetics and pathogenesis of polycystic kidney disease. J Am Soc Nephrol. 2002;13(9):2384–98.
27. Park JH, Li L, Cai Y, Hayashi T, Dong F, Maeda Y, Rubin C, Somlo S, Wu G. Cloning and characterization of the murine pkd2 promoter. Genomics. 2000;66(3):305–12.

Moisan *et al. BMC Genomics* (2018) 19:515

Page 12 of 13

28. Rodova M, Islam MR, Maser RL, Calvet JP. The polycystic kidney disease-1 promoter is a target of the beta-catenin/T-cell factor pathway. J Biol Chem. 2002;277(33):29577–83.

29. Lantinga-van Leeuwen IS, Leonhard WN, Dauwerse H, Baelde HJ, van Oost BA, Breuning MH, Peters DJ. Common regulatory elements in the polycystic kidney disease 1 and 2 promoter regions. Eur J Hum Genet. 2005;13(5):649–59.

30. Gresh L, Fischer E, Reimann A, Tanguy M, Garbay S, Shao X, Hiesberger T, Fiette L, Igarashi P, Yaniv M, et al. A transcriptional network in polycystic kidney disease. EMBO J. 2004;23(7):1657–68.

31. Wang Q, Han G, Ye J, Gao X, Niu H, Zhao J, Chai Y, Li N, Yin H. Characterization of the polycystic kidney disease 2 gene promoter. Genomics. 2014;104(6 Pt B):512–9.

32. Li Q, Montalbetti N, Shen PY, Dai XQ, Cheeseman CI, Karpinski E, Wu G, Cantiello HF, Chen XZ. Alpha-actinin associates with polycystin-2 and regulates its channel activity. Hum Mol Genet. 2005;14(12):1587–603.

33. Li X, Luo Y, Starremans PG, McNamara CA, Pei Y, Zhou J. Polycystin-1 and polycystin-2 regulate the cell cycle through the helix-loop-helix inhibitor Id2. Nat Cell Biol. 2005;7(12):1202–12.

34. Li Y, Wright JM, Qian F, Germino GG, Guggino WB. Polycystin 2 interacts with type I inositol 1,4,5-trisphosphate receptor to modulate intracellular Ca2+ signaling. J Biol Chem. 2005;280(50):41298–306.

35. Ma R, Li WP, Rundle D, Kong J, Akbarali HI, Tsiokas L. PKD2 functions as an epidermal growth factor-activated plasma membrane channel. Mol Cell Biol. 2005;25(18):8285–98.

36. Rundle DR, Gorbsky G, Tsiokas L. PKD2 interacts and co-localizes with mDia1 to mitotic spindles of dividing cells: role of mDia1 IN PKD2 localization to mitotic spindles. J Biol Chem. 2004;279(28):29728–39.

37. Tsiokas L, Arnould T, Zhu C, Kim E, Walz G, Sukhatme VP. Specific association of the gene product of PKD2 with the TRPC1 channel. Proc Natl Acad Sci U S A. 1999;96(7):3934–9.

38. Tsiokas L, Kim E, Arnould T, Sukhatme VP, Walz G. Homo- and heterodimeric interactions between the gene products of PKD1 and PKD2. Proc Natl Acad Sci U S A. 1997;94(13):6965–70.

39. Tran U, Zakin L, Schweickert A, Agrawal R, Doger R, Blum M, De Robertis EM, Wessely O. The RNA-binding protein bicaudal C regulates polycystin 2 in the kidney by antagonizing miR-17 activity. Development. 2010;137(7):1107–16.

40. Bouvrette DJ, Sittaramane V, Heidel JR, Chandrasekhar A, Bryda EC. Knockdown of bicaudal C in zebrafish (Danio rerio) causes cystic kidneys: a nonmammalian model of polycystic kidney disease. Comp Med. 2010;60(2):96–106.

41. Yang J, Zheng W, Wang Q, Lara C, Hussein S, Chen XZ. Translational up-regulation of polycystic kidney disease protein PKD2 by endoplasmic reticulum stress. FASEB J. 2013;27(12):4998–5009.

42. Zheng W, Shen F, Hu R, Roy B, Yang J, Wang Q, Zhang F, King JC, Sergi C, Liu SM, et al. Far upstream element-binding protein 1 binds the 3' untranslated region of PKD2 and suppresses its translation. J Am Soc Nephrol. 2016;27(9):2645–57.

43. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002;295(5558):1306–11.

44. Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C, et al. Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res. 2006;16(10):1299–309.

45. Moisan S, Berlivet S, Ka C, Gac GL, Dostie J, Ferec C. Analysis of long-range interactions in primary human cells identifies cooperative CFTR regulatory elements. Nucleic Acids Res. 2016;44(6):2564–76.

46. Trujillano D, Bullich G, Ossowski S, Ballarin J, Torra R, Estivill X, Ars E. Diagnosis of autosomal dominant polycystic kidney disease using efficient PKD1 and PKD2 targeted next-generation sequencing. Mol Genet Genomic Med. 2014;2(5):412–21.

47. Bogdanova N, Markoff A, Gerke V, McCluskey M, Horst J, Dworniczak B. Homologues to the first gene for autosomal dominant polycystic kidney disease are pseudogenes. Genomics. 2001;74(3):333–41.

48. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485(7398):376–80.

49. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159(7):1665–80.

50. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W, et al. Chromatin architecture reorganization during stem cell differentiation. Nature. 2015;518(7539):331–6.

51. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326(5950):289–93.

52. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al. Spatial partitioning of the regulatory landscape of the X-inactivation Centre. Nature. 2012;485(7398):381–5.

53. Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat Rev Genet. 2013;14(6):390–403.

54. Dixon JR, Gorkin DU, Ren B. Chromatin domains: the unit of chromosome organization. Mol Cell. 2016;62(5):668–80.

55. Smith EM, Lajoie BR, Jain G, Dekker J. Invariant TAD boundaries constrain cell-type-specific looping interactions between promoters and distal elements around the CFTR locus. Am J Hum Genet. 2016;98(1):185–201.

56. Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, Ong CT, Hookway TA, Guo C, Sun Y, et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell. 2013;153(6):1281–95.

57. Fraser J, Ethier SD, Miura H, Dostie J. A torrent of data: mapping chromatin organization using 5C and high-throughput sequencing. Methods Enzymol. 2012;513:113–41.

58. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature. 2012;489(7414):109–13.

59. Berlivet S, Paquette D, Dumouchel A, Langlais D, Dostie J, Kmita M. Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs. PLoS Genet. 2013;9(12):e1004018.

60. Rousseau M, Crutchley JL, Miura H, Suderman M, Blanchette M, Dostie J. Hox in motion: tracking HoxA cluster conformation during differentiation. Nucleic Acids Res. 2014;42(3):1524–40.

61. Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet. 2014;15(4):234–46.

62. Hnisz D, Day DS, Young RA. Insulated neighborhoods: structural and functional units of mammalian gene control. Cell. 2016;167(5):1188–200.

63. Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CW, Ye C, Ping JL, Mulawadi F, et al. CTCF-mediated functional chromatin interactome in pluripotent cells. Nat Genet. 2011;43(7):630–8.

64. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell. 2007;128(6):1231–45.

65. Hopp K, Ward CJ, Hommerding CJ, Nasr SH, Tuan HF, Gainullin VG, Rossetti S, Torres VE, Harris PC. Functional polycystin-1 dosage governs autosomal dominant polycystic kidney disease severity. J Clin Invest. 2012;122(11):4257–73.

66. Bhattacharjee S, Renganaath K, Mehrotra R, Mehrotra S. Combinatorial control of gene expression. Biomed Res Int. 2013;2013:407263.

67. Teng L, He B, Gao P, Gao L, Tan K. Discover context-specific combinatorial transcription factor interactions by integrating diverse ChIP-Seq data sets. Nucleic Acids Res. 2014;42(4):e24.

68. Leonhard WN, Happe H, Peters DJ. Variable cyst development in autosomal dominant polycystic kidney disease: the biologic context. J Am Soc Nephrol. 2016;27(12):3530–8.

69. Lupianez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015;161(5):1012–25.

70. Vecchio-Pagan B, Blackman SM, Lee M, Atalar M, Pellicore MJ, Pace RG, Franca AL, Raraigh KS, Sharma N, Knowles MR, et al. Deep resequencing of CFTR in 762 F508del homozygotes reveals clusters of non-coding variants associated with cystic fibrosis disease traits. Human Genome Variation. 2016;3:16038.

71. Mallawaarachchi AC, Hort Y, Cowley MJ, McCabe MJ, Minoche A, Dinger ME, Shine J, Furlong TJ. Whole-genome sequencing overcomes pseudogene homology to diagnose autosomal dominant polycystic kidney disease. Eur J Hum Genet. 2016;24(11):1584–90.

72. Hu Q, Wu Y, Tang J, Zheng W, Wang Q, Nahirney D, Duszyk M, Wang S, Tu JC, Chen XZ. Expression of polycystins and fibrocystin on primary cilia of lung cells. Biochem Cell Biol. 2014;92(6):547–54.

73. Dostie J, Dekker J. Mapping networks of physical interactions between genomic elements using 5C technology. Nat Protoc. 2007;2(4):988–1002.

74. Ferraiuolo MA, Rousseau M, Miyamoto C, Shenker S, Wang XQ, Nadler M, Blanchette M, Dostie J. The three-dimensional architecture of Hox cluster silencing. Nucleic Acids Res. 2010;38(21):7472–84.

Moisan *et al. BMC Genomics* (2018) 19:515

Page 13 of 13

75. Lajoie BR, van Berkum NL, Sanyal A, Dekker J. My5C: web tools for chromosome conformation capture studies. Nat Methods. 2009;6(10):690–1.

76. Fraser J, Rousseau M, Blanchette M, Dostie J. Computing chromosome conformation. Methods Mol Biol. 2010;674:251–68.

77. Dostie J, Zhan Y, Dekker J: Chromosome conformation capture carbon copy technology. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] 2007, Chapter 21:Unit 21 14.

78. Rousseau M, Ferraiuolo MA, Crutchley JL, Wang XQ, Miura H, Blanchette M, Dostie J. Classifying leukemia types with chromatin conformation data. Genome Biol. 2014;15(4):R60.

79. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002;12(6):996–1006.