Ass

Christel Krueger[1*], Michelle R. King[1], Felix Krueger[2], Miguel R. Branco[3¤a¤b], Cameron S. Osborne[4], Kathy K. Niakan[5,6], Michael J. Higgins[7], Wolf Reik[1,5*]

1 Epigenetics Programme, The Babraham Institute, Cambridge, United Kingdom, 2 Bioinformatics Group, The Babraham Institute, Cambridge, United Kingdom, 3 Genome Function Group, MRC Clinical Sciences Centre, Imperial College School of Medicine, Hammersmith Hospital Campus, London, United Kingdom,

interactions between defined genomic elements, or alternatively may be caused by the properties of large chromosomal regions. In this study, we have investigated in detail the pairing properties of various imprinted and non-imprinted genomic regions, and explored the possibility that pairing occurs for a specific purpose outside specialised settings such as X-inactivation or allelic exclusion. We find pairing frequency to be dependent on chromosomal position and transcriptional activity, rather than on mono- or biallelic expression of genes. We propose thatbybyvar29ve pai285dftar293los7.1adai285dthan

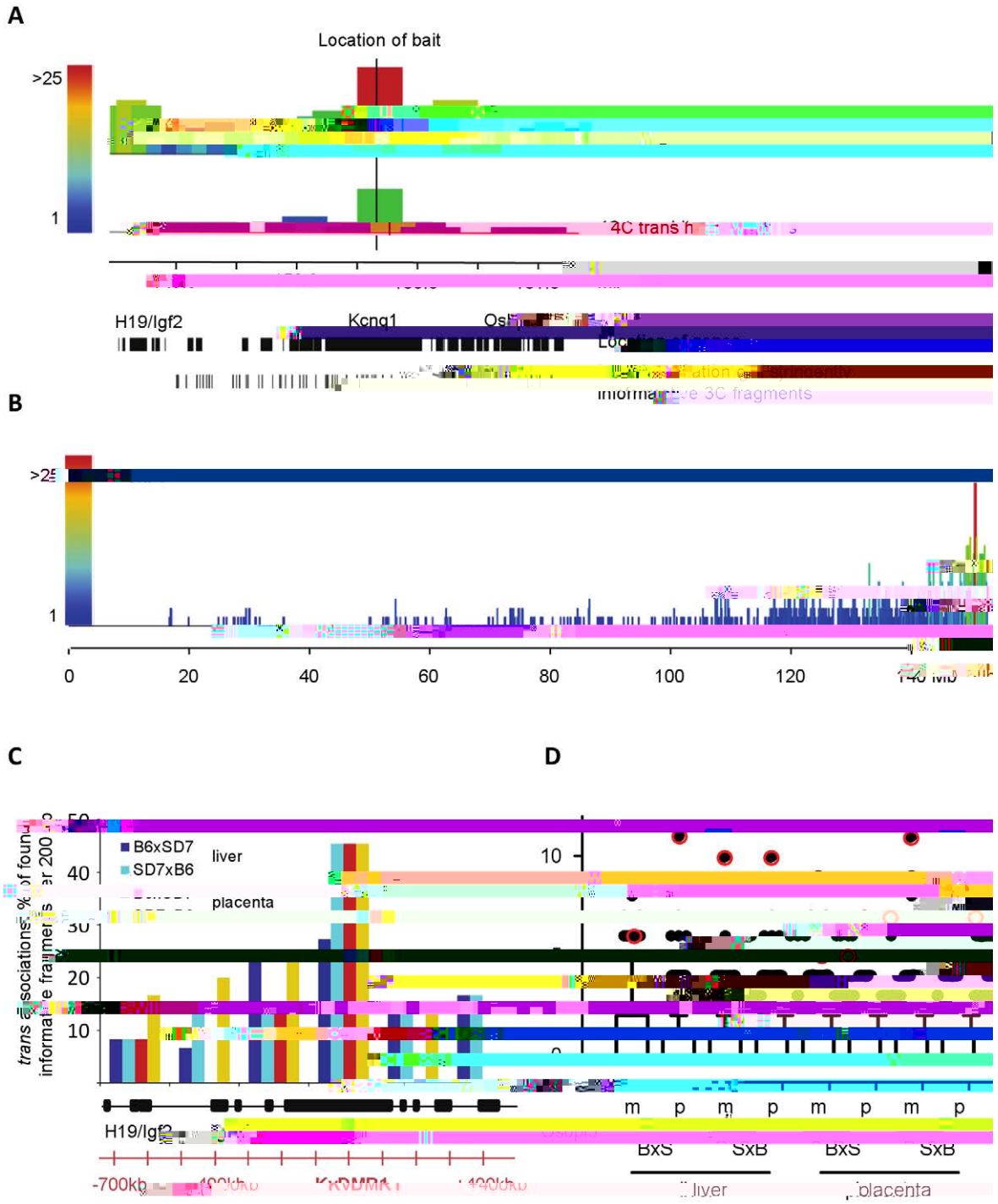

Figure 1. 4C-Seq reveals *trans*-allelic associations. A) Example of 4C-Seq association profile surrounding the KvDMR bait in the middle of the *Kcnq1* gene (window size 1.5 Mb, sample B6xSD7, E13.5 liver). The first row shows the quantification of all non-duplicated 4C-Seq reads per 100 kb window for the maternal bait in *cis*. 4C-Seq reads can only occur at certain positions and each position was counted only once (see File S2 for details). Colour and height of the bar reflect how many positions were found per 100 kb window (31 for the window that includes the bait). The second row shows associations of the maternal bait (B6) with the *trans* (paternal, SD7) allele. All reads identified as SD7 specific by ASAP were quantified per 100 kb window (see File S2 for details), again counting every position only once. A scale bar indicates the location of the region on chromosome 7. Black bars below represent the location of genes with some labelled for orientation. The positions of 3C fragments classified as 'stringently informative' are depicted at the bottom (see File S2 for details). The vast majority of associations occurs in *cis*. *Trans*-allelic associations occur most frequently with the homologous region. B

somal regions [21]. We therefore assessed if pairing was linked with the expression state of a region. At the *Kcnq1* locus, the paternal allele is transcriptionally silenced while a number of protein coding genes are expressed from the maternal allele. It has been shown by RNA immuno FISH that the active maternal allele colocalises with regions of high RNA polymerase II (RNA PolII) concentration while the paternal allele does not [17]. Accordingly, by DNA immuno FISH we generally find only one allele covered by signal for active RNA PolII (Fig. 5). Strikingly, when KvDMR FISH signals were paired, they both colocalised with RNA PolII demonstrating that pairing events occur in regions of active transcription (Fig. 5). We then assessed a link between pairing frequency of regions analysed by DNA FISH and their expression level (published RNA-Seq data, [31]). Overall, we found a significant correlation between expression and pairing frequency ($r = 0.62$, $p = 0.01$, Fig. S6A) but not between gene density and pairing frequency ($r = 0.30$, $p = 0.26$, Fig. S6B) indicating that active transcription is a key factor for pairing. Interestingly, the -2 Mb probe which shows the highest pairing frequency in the analysis lies within a gene dense region that is highly transcribed, but not known to be monoallelically expressed (Fig. S6).

Discussion

We have demonstrated for the first time by allele specific 4C-Seq and by extensive DNA FISH analysis that many loci pair with their homologous allele. Pairing is not limited to regions of mono-allelic expression, involves larger chromosomal regions and brings the two homologous chromosomes into close proximity. While pairing events did not coincide with DNA repair, they took place

there is no immediate requirement for communication between homologous alleles. Nevertheless, short interallelic distances were observed in late S phase for the Prader-Willi/Angelman imprinted region in human [5], although this effect was argued by others to be due to the presence of a nucleolus organising region on the same chromosome [33]. Here we report high pairing frequency for the *Kcnq1* and adjacent *Ins2/H19* clusters in the mouse, but not for a number of other imprinted clusters. Pairing at distal chromosome 7 was not limited to the imprinted region, and in fact loss of imprinting did not change pairing frequency. Thus, we conclude that homologous pairing is not a general feature of monoallelically expressed regions. However, this does not preclude an involvement of pairing and [illegible]-allelic effects on the regulation of imprinted regions. In fact, it was shown that introducing a third copy of human chromosome 15 disrupted pairing and affected gene expression at the Prader-Willi/Angelman region [34]. Speculatively, at the large *Kcnq1* domain which is silenced by a coating RNA, homologous pairing might be involved in the escape of imprinting of several interspersed biallelically expressed genes, especially as we find that pairing is associated with transcription.

Homologous pairing has also been speculated to be linked with DNA repair. The genome is constantly challenged by double strand breaks (DSB) brought about by internal or external chemical insults or the collapse of stalled replication forks (for review see [35]). These lesions can either be repaired by non-homologous end joining (NHEJ) or homologous recombination (HR). Which repair pathway is used depends on the organism and what caused the double strand break. While HR predominates in yeast, NHEJ plays a more important role in vertebrates. Still, in mammals HR is a common mechanism to repair replication induced damage after fork collapse which leaves a single double strand end. In this scenario the sister chromatid can be used as a template for strand invasion and restart of replication, a process that is helped by sister chromatid cohesion [36]. While it can be envisioned that more severe replication blocks may be repaired by HR involving both homologues, we did not find any evidence that links the homologous pairing described here with DSB or HR repair.

A number of recent genome-wide interaction studies in human cells have demonstrated the presence of topologically distinct active and repressive compartments, with [illegible] associations happening preferentially between transcriptionally active regions [37–39]. Moreover, a high frequency of [illegible] contacts correlated well with the region's distance to the edge of the chromosome territory [39]. As these studies were not performed in an allele-specific manner, no information about homologous contacts can be drawn. However, it seems likely that for a given region the criteria for a high potential of [illegible] interactions, namely

two spots were discernible by eye, ii) the MetaCyte signal annotation was in the middle of the two spots and iii) there were no other signals visible in the nucleus. Since the z-planes of the image stacks are 0.5 µm apart, this was considered the maximal resolution of the analysis. Paired signals are therefore less than 0.5 µm apart.

For DNA immuno FISH, pairing events were identified using automated image capture followed by manual acquisition of image stacks using ISIS software (Metasystems, version 5.4). For chromosome painting, nuclei were imaged on an Olympus IX81 confocal microscope (FV1000) using a 60×1.35 NA plan apochromat lens and Olympus fluoview software (version 3.0). Deconvolution of captured image stacks was performed with Huygens Professional software (Scientific Volume Imaging, version 4.1). Imaris software (Bitplane, version 7.3) was used for image analysis and 3D modelling.

Simulated FISH

A computational model was developed in R to simulate the position of two FISH signals inside the nucleus which display preferential radial positions. The constraint on radial distribution is attained by setting up two exclusion limits for each allele in the nucleus: a central exclusion limit (minimum distance from the nuclear centre) and a peripheric exclusion limit (maximum distance from the centre). For each simulated signal, the limits are randomly chosen based on a normal distribution whose mean and standard deviation are input by the user. By adjusting the input variables the radial distribution of simulated signals is matched to the radial distribution of the respective measured FISH signals. As an output, the model displays the distance between two simulated FISH signals.

Enrichment for Cell Cycle Stages

ES cells were fixed in 4% PFA in PBS for 10 min. Dye Cycle Violet staining was performed according to instructions of the manufacturer (Vybrant Dye Cycle Violet stain, Molecular Probes). Briefly, fixed cells were washed in 0.1 M Tris-Cl (pH 7.4) and then incubated with 10 µM Dye Cycle Violet in 0.1 M Tris-Cl for 30 min at 37°C. Samples were FACS sorted on a BD FACS Aria3 using violet 405 nm excitation with a 450/40 nm bandpass filter. Cell fractions were attached to slides using a cytospin (300 rpm, 3 min). DNA FISH was performed as normal.

Supporting Information

Figure S1 Radial distributions of DNA FISH signals and simulated counterparts. Radial distances for KvDMR (A) and myc (B) DNA FISH signals from E13.5 liver nuclei were determined and plotted as a histogram (grey bars, n = 600). Radial distances >1 can occur if the nucleus is not a perfect sphere. For both probes, FISH signals show a highly non-random radial distribution and thus cannot be compared to a simulated data set in which signals are placed in a sphere at random. Therefore, locations of signals were simulated to reflect the radial distribution of the respective probe (blue line, n = 600), but are otherwise random. The distance between pairs of simulated signals was then calculated and is plotted in Fig. 2D. C) Histogram of interallelic distances of KvDMR and myc signals, measured and simulated. Interallelic distances for 600 nuclei were grouped into four equal bins (bin width = 0.7r). Bin centres are indicated (0, 0.7, 1.4 and 2.1r). The histogram displays the same data that is shown as Tukey box-whisker plots in Fig. 2D to illustrate the skewing towards very short distances for the measured KvDMR signals.
(TIF)

Figure S2 Pairing of the KvDMR region in ES cells. A) Pairing frequency of the KvDMR and myc regions in ES cells. Each dot represents one sample from one cell passage, and for each sample 300 nuclei were counted in four technical replicates. Differences were assessed by unpaired t-test, ns: not significant, *: $p<0.05$, ***: $p<0.001$. KvDMR signals show higher pairing frequency than myc signals. B) Distances between homologous alleles in E13.5 liver represented as Tukey box-whisker plots (n = 600). Differences were assessed by one-way ANOVA with Bonferroni's multiple comparison post test. Simulated: a group of spots displaying the same radial distributions as KvDMR and myc FISH signals, respectively, were placed into a sphere at random and their interallelic distances determined. While interallelic distances between myc signals show a distribution expected from their radial positions (no difference between sample and simulated), KvDMR signals are significantly closer together than expected (difference between sample and simulated, $p<0.001$). Also, KvDMR signals are generally closer together than myc signals.
(TIF)

Figure S3 Pairing frequencies of imprinted regions in ES cells. Each dot represents the mean of three to four samples taken from different passages with the pairing frequency determined for each sample in four technical replicates of 300 nuclei. Whiskers represent standard deviation. Background shading indicates high pairing frequency (dark grey, above 3.5%), medium (light grey, 2.5–3.5%) and low pairing frequency (yellow, below 2.5%).
(TIF)

Figure S4 Paired KvDMR FISH signals do not colocalise with markers for DNA double strand breaks or repair by homologous recombination. 3D representations of image stacks from E13.5 liver cells. Red: KvDMR DNA FISH signals, green: immunostaining for markers of DNA double strand breaks (γH2AX (A), p53bp1 (B)), or markers for homologous recombination (Rad51 (C), Rad52 (D)), blue: DAPI counterstain. Numbers for analysed pairing events are indicated. No overlap of immuno and FISH signals was observed.
(TIF)

Figure S5 Pairing is not specific to a certain cell cycle stage. A) Histogram of PFA fixed ES cells stained with DyeCycle

Table S2 Antibodies.
(DOC)

File S1 Derivation and characterisation of the ES cell line B6xSD7.
(DOC)

File S2 Details of the linear 4C-Seq analysis.
(DOC)

Movie S1 Regional pairing brings chromosome domains close together.
(MOV)

Acknowledgments

We would like to thank Daniel Bolland and Simon Walker for help with DNA FISH and imaging. We are grateful to Wendy Dean for help and advice, Geoff Morgan and Arthur Davis for FACS sorting, and Anne Segonds-Pichon for statistical input. We thank Matthew Turley for help with the characterisation of the ES cell line.

Author Contributions

Performed bioinformatic analysis: FK. Developed allele-specific 4C-Seq: CK CSO. Developed computational model: MRB. Directed the study: WR. Conceived and designed the experiments: CK MRK. Performed the experiments: CK MRK. Analyzed the data: CK MRK. Contributed reagents/materials/analysis tools: KKN MJH. Wrote the paper: CK.

References

1. Schoenfelder S, Clay I, Fraser P (2010) The transcriptional interactome: gene expression in 3D. Curr Opin Genet Dev 20: 127–133.
2. Duncan IW (2002) Transvection effects in Drosophila. Annu Rev Genet 36: 521–556.
3. Anguera MC, Sun BK, Xu N, Lee JT (2006) X-chromosome kiss and tell: how the Xs go their separate ways. Cold Spring Harb Symp Quant Biol 71: 429–437.
4. Hewitt SL, Yin B, Ji Y, Chaumeil J, Marszalek K, et al. (2009) RAG-1 and ATM coordinate monoallelic recombination and nuclear positioning of immunoglobulin loci. Nat Immunol 10: 655–664.
5. LaSalle JM, Lalande M (1996) Homologous association of oppositely imprinted chromosomal domains. Science 272: 725–728.
6. Thatcher KN, Peddada S, Yasui DH, Lasalle JM (2005) Homologous pairing of 15q11–13 imprinted domains in brain is developmentally regulated but deficient in Rett and autism samples. Hum Mol Genet 14: 785–797.
7. Hu JF, Vu TH, Hoffman AR (1997) Genomic deletion of an imprint maintenance element abolishes imprinting of both insulin-like growth factor II and H19. J Biol Chem 272: 20715–20720.
8. Duvillie B, Bucchini D, Tang T, Jami J, Paldi A (1998) Imprinting at the mouse Ins2 locus: evidence for cis- and trans-allelic interactions. Genomics 47: 52–57.
9. Koeman JM, Russell RC, Tan MH, Petillo D, Westphal M, et al. (2008) Somatic pairing of chromosome 19 in renal oncocytoma is associated with deregulated EGLN2-mediated [corrected] oxygen-sensing response. PLoS Genet 4: e1000176.
10. Stout K, van der Maarel S, Frants RR, Padberg GW, Ropers HH, et al. (1999) Somatic pairing between subtelomeric chromosome regions: implications for human genetic disease? Chromosome Res 7: 323–329.
11. Cremer M, von Hase J, Volm T, Brero A, Kreth G, et al. (2001) Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. Chromosome Res 9: 541–567.
12. Kozubek S, Lukasova E, Jirsova P, Koutna I, Kozubek M, et al. (2002) 3D Structure of the human genome: order in randomness. Chromosoma 111: 321–331.
13. Mayer R, Brero A, von Hase J, Schroeder T, Cremer T, et al. (2005) Common themes and cell type specific variations of higher order chromatin arrangements in the mouse. BMC Cell Biol 6: 44.
14. Heride C, Ricoul M, Kieu K, von Hase J, Guillemot V, et al. (2010) Distance between homologous chromosomes results from chromosome positioning constraints. J Cell Sci 123: 4063–4075.
15. Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, et al. (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell 32: 232–246.
16. Redrup L, Branco MR, Perdeaux ER, Krueger C, Lewis A, etJ, 34l(rB(L,)e8C8(eehem)]TJ1.967-1.138Td[loel)-3t.o7s alMR,5(arrange)-18s5i4tissl at (2008)-269(Polycombn)-24(ghroup)278(pnroeions)-285.6Exh2pf2p genofic co(trceti)-11.4(n)-2690.7(and)]TJ0-1.138TD[3(imirited,)-34326(rprnesivon)-374.9(in)-332.7eaelyn mousl mbryosl. Ce
86. Lewis A, K, C, Redrup Huynh. KD,o 2063–2410.
A,96. Lewis A, P, et al. (204(L,)e2364(Iimirit.3(alM2-9.3mon)]TJ1.967-1.146Td[dDistal)-715.8(chrom)-9.8mosose
5. nide of
uf.YA,acoRR, Crqueitra F, Wags(ch)-1259allet al. (204(L,)e334.6Iimirit.ing
alM616.2(the)-686.8966(5e)-73063dsom(in)-674.5oin
nadrcruit(men,)5049.6(of)-490Polycombn ghroupcmplexesNat. 6:
S,Sextmon T,va
tsose inthridl
rmeder P, vanimone

45. Forne T, Oswald J, Dean W, Saam JR, Bailleul B, et al. (1997) Loss of the maternal H19 gene induces changes in Igf2 methylation in both cis and trans. Proc Natl Acad Sci U S A 94: 10243–10248.
46. Herman H, Lu M, Anggraini M, Sikora A, Chang Y, et al. (2003) Trans allele