

ORIGINAL ARTICLE

High-sensitivity sequencing reveals multi-organ somatic mosaicism causing DICER1 syndrome

Leanne de Kock,^{1,2} Yu Chang Wang,³ Timothée Revil,³ Dunarel Badescu,³ Barbara Rivera,^{1,2} Nelly Sabbaghian,² Mona Wu,^{1,2} Evan Weber,⁴ Claudio Sandoval,⁵ Saskia M J Hopman,⁶ Johannes H M Merks,⁶ Johanna M van Hagen,⁷ Antonia H M Bouts,⁸ David A Plager,⁹ Aparna Ramasubramanian,^{9,10} Linus Forsmark,¹¹ Kristine L Doyle,¹² Tonja Toler,¹³ Janine Callahan,¹⁴ Charlotte Engelenberg,¹⁵ Dorothée Bouron-Dal Soglio,¹⁶ John R Priest,¹⁷ Jiannis Ragoussis,³ William D Foulkes^{1,2,4,18}

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jmedgenet-2015-103428>)

For numbered affiliations see end of article.

Correspondence to

Dr William D Foulkes, Department of Medical Genetics, Lady Davis Institute, Segal Cancer Centre, Jewish General Hospital, 3755 Cote St. Catherine Road, Montreal, Québec, Canada H3T 1E2; william.foulkes@mcgill.ca

Received 30 July 2015

Revised 7 September 2015

Accepted 18 September 2015

Published Online First

16 October 2015

ABSTRACT

Background Somatic mosaicism is being increasingly recognised as an important cause of non-Mendelian presentations of hereditary syndromes. A previous whole-exome sequencing study using DNA derived from peripheral blood identified mosaic mutations in *DICER1* in two children with overgrowth and developmental delay as well as more typical phenotypes of germline *DICER1* mutation. However, very-low-frequency mosaicism is difficult to detect, and thus, causal mutations can go unnoticed. Highly sensitive, cost-effective approaches are needed to molecularly diagnose these persons. We studied four children with multiple primary tumours known to be associated with the DICER1 syndrome, but in whom germline *DICER1* mutations were not detected by conventional mutation detection techniques.

Methods and results We observed the same missense mutation within the *DICER1* RNase IIIb domain in multiple tumours from different sites in each patient, raising suspicion of somatic mosaicism. We implemented three different targeted-capture technologies, including the novel HaloPlex^{HS} (Agilent Technologies), followed by deep sequencing, and confirmed that the identified mutations are mosaic in origin in three patients, detectable in 0.24–31% of sequencing reads in constitutional DNA. The mosaic origin of patient 4's mutation remains to be unequivocally established. We also discovered likely pathogenic second somatic mutations or loss of heterozygosity (LOH) in tumours from all four patients.

Conclusions Mosaic *DICER1* mutations are an important cause of the DICER1 syndrome in patients with severe phenotypes and often appear to be accompanied by second somatic truncating mutations or LOH in the associated tumours. Furthermore, the molecular barcode-containing HaloPlex^{HS} provides the sensitivity required for detection of such low-level mosaic mutations and could have general applicability.

INTRODUCTION

Mosaicism arises following the acquisition of a de novo mutation during post-zygotic development, resulting in an individual with two populations of cells that are genetically distinct.¹ Mosaicism is

being increasingly recognised as the cause of a diverse range of sporadic albeit likely genetic clinical disorders, the aetiology of which was previously unknown.^{1–4} This is largely attributed to improved genomic sequencing technologies that have provided better ability to detect genetic changes in subpopulations of cells. Despite this, detecting low-level mosaicism is still challenging.

The DICER1 syndrome or pleuropulmonary blastoma (PPB) familial tumour and dysplasia syndrome (OMIM #601200) is typically caused by heterozygous germline mutations in *DICER1*, which encodes a small RNA endoribonuclease responsible for processing hairpin precursor miRNAs into mature miRNAs that in turn post-transcriptionally regulate expression of target messenger RNAs. Since studies began on *DICER1* in 2009,⁵ >140 heterozygous germline mutations have been published.⁶ DICER1 syndrome is associated with a predisposition to several rare phenotypes including PPB, cystic nephroma (CN), ovarian Sertoli–Leydig cell tumour (SLCT), multinodular goitre (MNG), nasal chondromesenchymal hamartoma (NCMH), pineoblastoma, pituitary blastoma (PitB) and other rare conditions.⁶ DICER1 syndrome has autosomal-dominant inheritance with variable penetrance and may present from infancy through adolescence and occasionally later.⁶ Using whole-exome sequencing, Klein *et al*⁷ have recently reported mosaic missense mutations in *DICER1*, affecting the RNase IIIb metal-ion binding domain, in peripheral blood DNA from two infants; allele frequencies were 21% and 28%, respectively. Each infant had extensive bilateral lung cysts consistent radiographically with cystic PPB, developed bilateral Wilms tumour (in one child in large kidneys with underlying renal dysmorphology) and had global developmental delays and various overgrowth stigmata. Klein *et al* used 'Global delay, Lung cysts, Overgrowth and Wilms (GLOW)' syndrome to describe this phenotype. Brief mention of mosaic *DICER1* mutations also appears in one abstract.⁸

Here, we describe our detailed molecular investigation of four children with multiple primary tumours consistent with the DICER1 syndrome



CrossMark

To cite: de Kock L, Wang YC, Revil T, *et al*. *J Med Genet* 2016;**53**:43–52.

phenotype, but in whom germline *DICER1* mutations had not been detected by Sanger sequencing or multiplex ligation-based probe amplification (MLPA) assay⁹ of genomic DNA isolated from peripheral blood or saliva.

MATERIALS AND METHODS

Details of sample acquisition, gDNA extractions, Sanger sequencing and the MLPA assay are presented in the online supplementary material methods.

Targeted captures and next-generation sequencing

We interrogated both tumour and normal tissues for evidence of somatic mosaicism using a standard HaloPlex targeted capture¹⁰ (Agilent Technologies, Santa Clara, USA), an in-house Fluidigm Access Array (Fluidigm, San Francisco, California, USA), and a novel development of the HaloPlex assay that incorporates molecular barcodes for high-sensitivity sequencing as a custom design (HaloPlex^{HS}).

In brief, the HaloPlex^{HS} targeted capture method is specifically designed to identify low allele frequency variants through

the attachment of a 10-nucleotide-long molecular barcode to the captured sample DNA molecules. High sensitivity is supported by the capture of up to eight different restriction fragments per targeted base in the region of interest (figure 1). In our case, >75% of the targeted bases were covered by at least four probes (see online supplementary figures S1A-B and S2). During downstream analysis of the sequencing data, molecular barcode sequence data are used to collapse reads originating from the same sample molecule, which improves base calling accuracy and allows for accurate quantification of the mutant allele fraction within each sample as it excludes possible PCR amplification bias. The design used in this study captures 499 kb and encompasses the full *DICER1*, *DROSHA*, *AGO1*, *TRBP2* and *DGCR8* loci, all miRNA-processing-associated genes (see online supplementary figure S1).^{6–11} We will be pleased to make the design of this array available; please contact the authors for further details. We processed three gDNA samples from different sites from each child for a total of 12 samples (excluding controls), 11 of which were non-tumourous. The sequencing data were generated on the Illumina HiSeq2500 sequencer

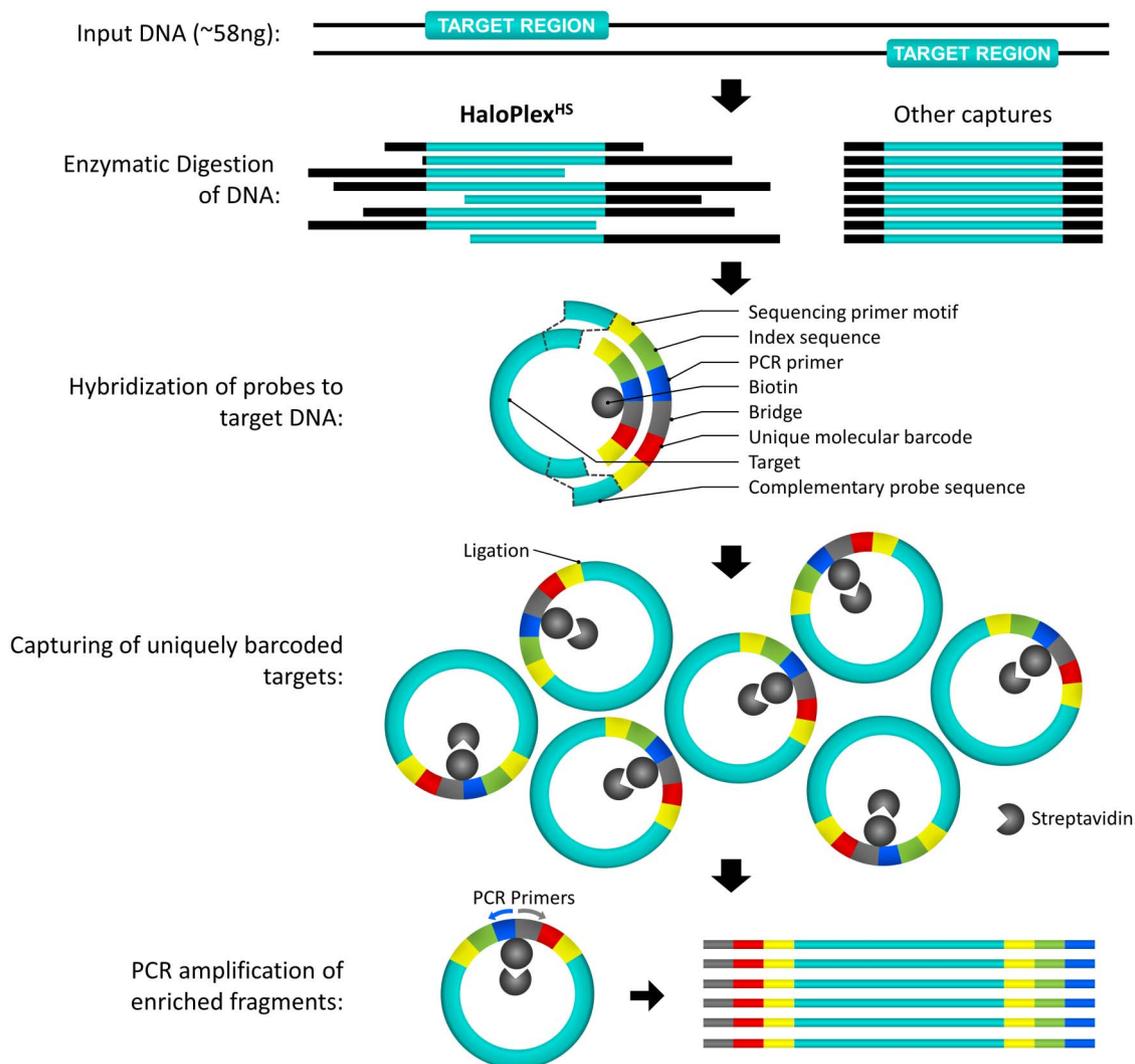


Figure 1 Graphical representation of the HaloPlex^{HS} design principle. The HaloPlex^{HS} captures up to eight different restriction fragments per targeted base in the region of interest, ensuring that the great majority of target region is ultimately covered. During hybridisation, each sample is given a unique index sequence (green), allowing for pooling of up to 96 samples per sequencing lane. A degenerate molecular barcode sequence (red) is also incorporated during hybridisation, which makes it possible to track individual target amplicons during sequence analysis and to remove duplicate reads if necessary.

using the 150 bp paired-end sequencing protocol across four rapid flow-cell lanes. The depth of coverage achieved for each sample is depicted in online supplementary figure S1C and D. We also used the standard HaloPlex targeted capture system with a similar probe design (see online supplementary figure S2), but which does not incorporate molecular barcodes to facilitate the removal of duplicate reads. We sequenced a total of 28 gDNA samples using the Illumina MiSeq sequencer, 17 of which were non-tumourous and 11 were tumour derived. Lastly, we used a custom design Fluidigm Access Array that selectively captures all exons and exon–intron boundaries of *DICER1* (described previously)¹² to allow for cross-platform validation of our findings. Following capture with Fluidigm, we sequenced 22 normal gDNA samples and 15 tumour gDNA samples using the Illumina MiSeq for a total of 37 samples (see online supplementary figure S3).

Inspection of suspected mosaic mutations

Initial analysis of the suspected mosaic missense mutations in the HaloPlex^{HS} data was performed using the SureCall software V3.0 (beta version), provided by Agilent Technologies. Sequencing reads were aligned to the human genome (hg19) and the duplicate reads were removed. The relative frequency of the four bases at the position of each mutation were manually inspected and recorded. We analysed the standard HaloPlex data in the same way using the then publically available SureCall software (V2.0.7.0) and analysed the Fluidigm-generated data using the Integrative Genomics Viewer software (IGV V2.3).

Calculation of threshold of detection in the HaloPlex^{HS} data

We calculated the percentage of false nucleotides at the position of the respective mosaic *DICER1* RNase IIIb mutations in each of the samples by dividing the number of reads containing an aberrant base by the total number of reads, multiplied by 100 (eg, blood DNA from patient 2 contained 4 of 9650 reads with a false T allele, or 0.04%). Using all samples with >100 reads covering the region of interest, we calculated the median number of false reads to be 0.04% (range 0–0.35%). By studying the distribution of false positive reads at the positions of the four RNase IIIb mutations, we were able to calculate the total number of false reads. We then took the fourth quintile of this distribution (0.06%) as the cut-off, below which we regarded all mutation calls as false positive. To be conservative, we considered the threshold of detection to be 0.1%.

Bioinformatic methods

The standard HaloPlex and Fluidigm-generated data sets were separately analysed using our custom bioinformatics pipeline as follows: raw paired-end reads were trimmed using Trimmomatic V0.33¹³ to a minimum length of 30 nucleotides. Illumina Truseq adapters were removed in palindrome mode. A minimum Phred quality score of 30 was required for the 3' end. Single end reads as well as paired-end reads failing previous minimum quality controls were discarded. Individual read groups were aligned, using bwa V0.7.12, with default parameters,¹⁴ to the UCSC hg19 reference human genome from Illumina iGenomes website.¹⁵ Aligned reads from multiple read groups belonging to the same sample were indexed, sorted and merged using Sambamba V0.5.4,¹⁶ a faster implementation of the Samtools algorithms.^{17 18}

Various quality control parameters were used, including depth of coverage, based on metrics collected for each sample using bedtools V2.24.0¹⁹ and aggregated using custom Python V2.7.9 codes. We applied GATK V3.3.0 base quality score recalibration

and indel realignment.²⁰ As amplification duplicates were not removed, we also added '–nt NONE' parameter to change the corresponding default down-sampling behaviour.

We performed SNP and INDEL discovery and genotyping across each cohort of samples simultaneously using standard hard filtering parameters according to GATK Best Practices recommendations.^{21 22} In addition to HaplotypeCaller algorithm, we also used UnifiedGenotyper in separate runs. Samtools' new multiallelic calling model (–m parameter), as implemented in bcftools V1.2, was also used.²³ All variants were annotated with functional prediction using snpEFF V3.6.²⁴ Additionally, functional annotation of variants found in two public databases (ie, NCBI dbSNP V.142²⁵ and dbNSFP V2.8^{26 27}) were added using SnpSift, as part of the same software package.²⁸

HaloPlex^{HS} data set was analysed using the SureCall software V3.0 (beta version). All variants obtained from all three data sets were inserted into a Gemini database,²⁹ aggregated and selected according to snpEFF predictions. Finally, they were manually validated against read alignments using IGV software V2.3.^{30 31}

Comparison of locus coverage and percentage of homozygosity

The average coverage of the five targeted loci was calculated using GATK's DepthOfCoverage tool (see online supplementary figure S4). The total number of SNPs per locus was extracted from a Gemini database loaded with only the HaplotypeCaller variants (see above for details). The percentage of homozygosity was then calculated using the number of homozygous alternate or reference SNPs for the *DICER1* locus or the other four targeted loci.

RESULTS

The clinical presentation of each child is outlined in table 1 and figure 2 with further details available in the online supplementary data and figures S5 and S6. The clinical cases of both patients 1³² and 2³³ have been previously described. The family history for all four children was unremarkable (see online supplementary figure S5).

In most individuals with clinical features suggestive of the *DICER1* syndrome, germline truncating mutations in *DICER1* are accompanied by specific somatic 'hotspot' missense mutations occurring within the sequence encoding the RNase IIIb domain.^{6 12 34} Our initial Sanger sequencing efforts did not identify causal germline mutations in *DICER1*, but we reasoned that if the tumours contained a somatic *DICER1* mutation, the disease presentation could be due to an occult germline mutation in *DICER1* or in a closely related gene. We thus undertook Sanger sequencing of gDNA extracted from multiple tumour samples from each patient to determine whether they harboured a somatic *DICER1* mutation. We observed the same missense mutation within the sequence encoding the *DICER1* RNase IIIb domain in multiple tumours from each patient (*patient 1*: c.5125G>A [p.D1709N], number of tumours sequenced (n) =8; *patient 2*: c.5437G>C [p.E1813Q], n=6; *patient 3*: c.5439G>C [p.E1813D], n=5; and *patient 4*: c.5425G>A [p.G1809R], n=2) (figure 3). There are a diverse range of missense mutations reported to occur at these particular hotspot locations,⁶ with 215 reports of such hotspot mutations in the literature, 115 of which were confirmed somatic and the great majority of the rest are presumed likely somatic. They mostly affect 1 of 11 nucleotides within the sequence encoding the *DICER1* RNase IIIb domain.⁶ We, therefore, thought it very

Table 1 Clinical summary

Patient	Age at pathological diagnosis	Disease
Patient 1	In utero*	Bilateral lung cysts
	Birth* [†] ; 5 days	Bilateral CCAM, revised later to type I PPB [†]
	Birth*	Scrotal web and testicular cyst
	9 months* [†] ; 1.7 years	Multiple small bowel polyps: enteritis cystica profunda [†]
	25 months	Type II PPB [†]
	25 months	PPB brain metastasis #1 [†]
	3.1 years	PPB brain metastasis #2 [†]
	8 years	NCMH
Patient 2	10.5 years	PPB brain metastasis #3 [†]
	10.9 years	PPB brain metastasis #4 [†]
	13 months* [†] ; 1.2 years	Hamartomatous bilateral renal cysts, NOST
	13 months* [†] ; 2.5 years	Benign multifocal bilateral lung cysts, NOST
	Birth* [†] ; 2 years	L ocular 'pre-phthisical changes': vascular mass and recurrent retinal detachments
	7.7 years	Pineoblastoma [†]
	10.6 years	Follicular variant of papillary thyroid carcinoma [†]
	13.4 years	L ovarian SLCT [†]
	15.1 years	Unusual hamartomatous nasal polyp, not otherwise categorised [†]
	15.7 years	R ovarian SLCT [†]
17.2 years	R teratoid CBME [†]	
Patient 3	10 days	R CCAM [†] , revised to type I PPB [†] ; L lung cysts
	6 months	Hamartomatous juvenile intestinal polyps [†]
	1 month	R renal medullary malformation with disorganised collecting system and dilated lymphatic vessels [†]
	1.1 years	L CN [†]
	6.8 years	NCMH
Patient 4	11 months* [†]	Bilateral multifocal lung cysts; L CCAM [†] ; revised to type I PPB [†]
	1.7 years	R infected pulmonary cyst [†]
	4.1 years	L residual type I PPB with polypoid type II PPB [†]

*Age at detection.

[†]Pathological diagnosis.

CBME, ciliary body medulloepithelioma; CCAM, congenital cystic adenomatoid malformation; CN, cystic nephroma; L, left; NCMH, nasal chondromesenchymal hamartoma; NOS, not otherwise specified; PPB, pleuropulmonary blastoma; R, right; SLCT, Sertoli–Leydig cell tumour.

unlikely that the same hotspot mutation would occur in all independent disease foci in each child, raising suspicion of somatic mosaicism for the identified *DICER1* mutations.

We confirmed a mosaic distribution of the respective mutations in patients 1–3 using the three targeted-capture platforms: the mutant allele fraction was significantly higher in tumour samples and was detected at low levels in multiple normal tissues from the three patients. The average mutant allele frequency in non-tumour samples ranged from 0% to 13.58% in normal tissues from patient 1 (12.77–60.80% in tumour samples); from 0% to 1.33% normal in tissues from patient 2 (0–99.41% in tumours) and from 0% to 31.73% in normal tissues from patient 3 (6.55–91.89% in tumours). The numbers of sequencing reads and the percentage of mutant and wildtype alleles per patient sample are indicated in table 2, online supplementary table S1 and figure 4A–D. Comparing the four technologies used, it is clear that the HaloPlex^{HS} data offers greater precision, given the ability to remove duplicate reads and in so doing increase the base calling accuracy, as described above. We considered 0.1% to be the threshold for detection, below which all mutant alleles detected were considered false positive (see table 2 and the 'Materials and methods'). Of particular interest is patient 2, c.5437G>C, where in saliva DNA,

using HaloPlex^{HS}, we identified 5 of 1972 reads with the C allele (0.25%), whereas only one read was seen for a T allele at this position (0.05%). This T allele is clearly a false positive read. By contrast, in blood DNA, we identified 4 of 9650 reads (0.04%) for both the C and T alleles (table 2), suggesting that both the C and the T alleles are false positives. From this result, we conclude that the c.5437G>C mutation is present in saliva DNA, but not in blood DNA. The c.5437G>C mutation was also identified at low levels in urine (table 2).

We hypothesised that, in the setting of a mosaic *DICER1* RNase IIIb mutations, we might discover second somatic mutations outside of the RNase IIIb domain, which initiate two-hit tumorigenesis as seen in most *DICER1*-related tumours. All three data sets were analysed as described in the 'Materials and methods', an outline of which is presented in online supplementary figure S4. We identified individually distinct second somatic likely deleterious *DICER1* mutations in patient 2's left ovarian SLCT (c.4626_4626delG; p.Q1542Hfs*18) and sinonasal inflammatory polyp (c.4458_4458delA; p.K1486Nfs*4), in patient 3's NCMH (c.4651_4652insTGCT; p.E1551Vfs*7) and in patient 4's type II PPB, which arose in a pre-existing lung cyst (c.1966C>T; p.R656*) (see online supplementary figure S6A). Each of these second somatic mutations was validated via Sanger sequencing and is predicted to prematurely truncate the *DICER1* protein (see online supplementary table S1 and figure 4). Furthermore, we detected loss of heterozygosity (LOH) in one of patient 1's PPB brain metastases and in three additional lesions from patient 2 (follicular thyroid carcinoma, right SLCT and kidney cysts). Evidence of LOH in the tumours is supported by both quantitative and qualitative analyses, as presented in figure 4E–H, online supplementary figure S7 and table S2. In online supplementary figure S7A, LOH is evident in the tumours from patient 1, column 8, and in patient 2, columns 6 and 7, where there is a visible reduction in the average coverage of the *DICER1* locus relative to the respective germline samples. Similarly, in online supplementary figure S7B, there is a visible increase in the per cent SNP homozygosity for tumours occurring in patient 1, column 7, and patient 2, columns 5 and 6, which is indicative of LOH.

Despite the time interval between the cancer diagnosis and blood sampling, it is possible that we may be detecting traces of circulating tumour DNA in blood samples or infiltrating tumour cells in normal tissue sampled from areas adjacent to tumours. In addition to collecting tissues that are less likely to contain contaminating tumour DNA (eg, hair and saliva), we also wanted to carry out additional analyses to explicitly determine whether we were indeed picking up contaminating tumour DNA. Three of the above exonic somatic mutations were not detected at all in germline samples in the HaloPlex^{HS} data set. The remaining somatic mutation found in patient 4's type II PPB was detected at 0.04% in blood DNA, well below the 0.1% threshold for likely real mutations (see 'Materials and methods'), and we are thus able to establish that the mosaic mutation-containing alleles detected in the non-tumorous samples were not derived from infiltrating tumour cells or circulating tumour DNA. The identified second somatic mutations and the number of alternate and wildtype alleles per patient sample are summarised in online supplementary table S3.

DISCUSSION

With the use of a targeted approach, combined with deep and ultra-deep sequencing, we detected low-level *DICER1* mutant allele fractions in three patients exhibiting mosaicism for the detected mutations. The fourth case is also likely to be a mosaic

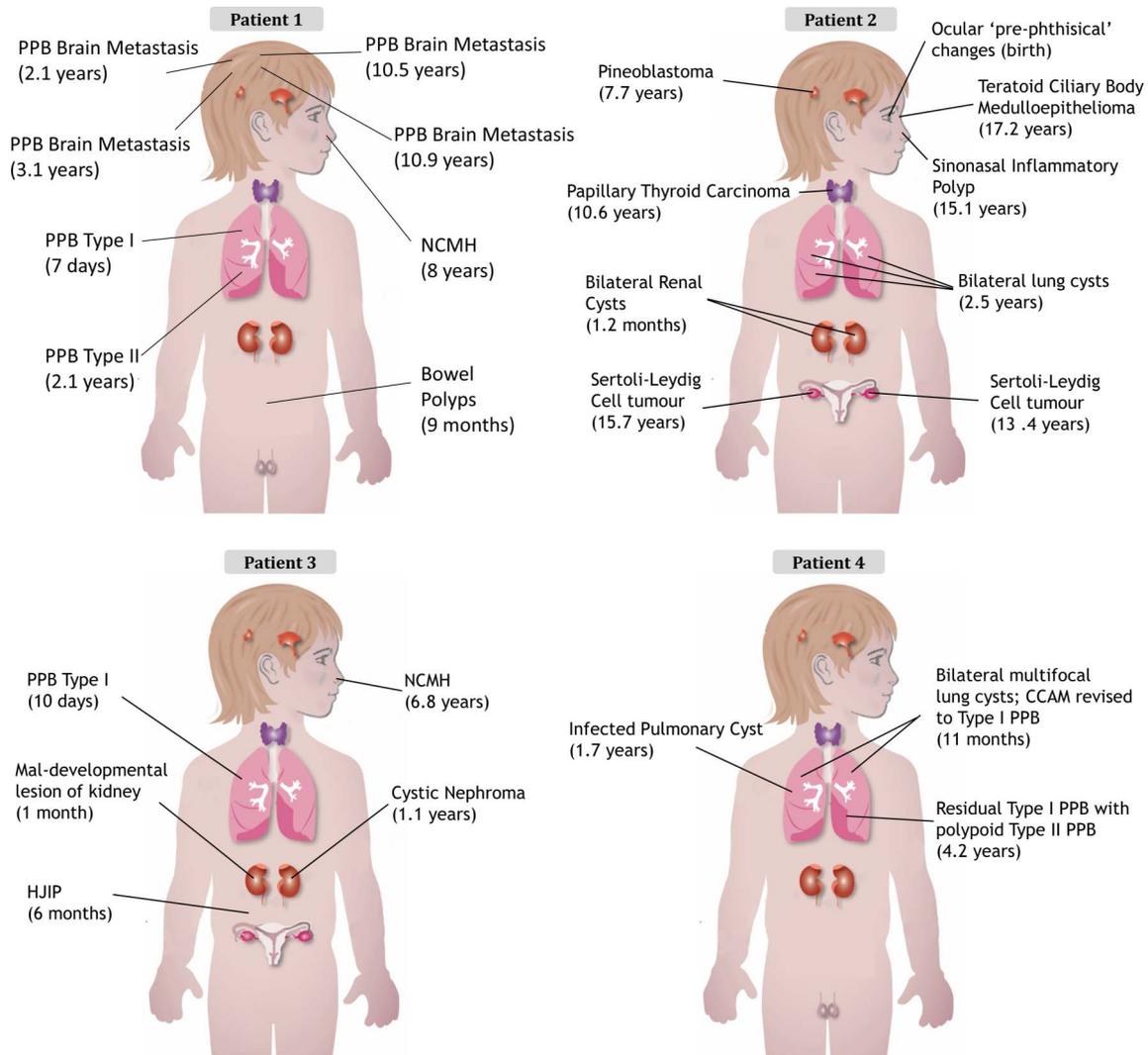


Figure 2 Diagnosis and (age at diagnosis) for each child's tumour. CCAM, congenital cystic adenomatoid malformation; HJIP, hamartomatous juvenile intestinal polyps; NCMH, nasal chondromesenchymal hamartoma; PPB, pleuropulmonary blastoma.

for *DICER1*. These mosaic missense mutations were localised to 'hotspots' within the sequence encoding the *DICER1* RNase IIIb domain and have been shown to selectively reduce 5p miRNA processing.^{34–36} We also discovered likely pathogenic second somatic mutations or LOH in tumours from all four patients, thus showing that the two-hit model applied to the tumours we studied (table 2 and online supplementary table S1).

The exact developmental stage at which the mosaic mutations were acquired has not been accurately determined, but given the presence of the mutant allele in tissue samples from all three germ layers, we suspect that the mutations occurred prior to gastrulation.^{1–37} The mosaic origin of patient 4's mutation remains to be unequivocally established (table 2, online supplementary table S1 and figures 3 and 4). Without additional normal samples from both the lung(s) and other distant normal sites, we were not able to determine whether (a) the child is a somatic mosaic with an undefined, yet limited distribution of the mutation; (b) mosaicism is present but is confined to the lungs; or (c) the two lung lesions separately acquired the c.5425G>A 'hit' by chance. In the latter case, the 6.9% mutant allele frequency in the reactive lung tissue would have to be attributed to cancer cells that were not obviously present on

detailed histopathological examination (see online supplementary figure S8), and therefore, we do not favour this explanation. Moreover, the detection of a second somatic truncating mutation in the PPB type II sample (see online supplementary table S1) and the absence of any further extrapulmonary *DICER1*-related lesions in this person support the hypothesis that somatic mosaicism is present but is confined to the lungs. In this case, acquisition of the missense mutation would have occurred much later during embryonic development than in the other three cases.

Of note, mosaic *DICER1* mutations in our cases and the two previously described cases⁷ are localised to the sequence encoding the RNase IIIb domain. We have also identified additional likely pathogenic second somatic mutations or LOH in the tumours. These findings strongly suggest that the molecular paradigm of multi-organ mosaic RNase IIIb mutations followed by second 'hits' in other regions of *DICER1* is precisely the reverse of typical, now well-described *DICER1* molecular events in which somatic RNase IIIb mutations follow inactivating germline mutations.⁶ This mosaic paradigm affecting the highly critical RNase IIIb residues may explain the apparently severe phenotype seen in three of our cases as well as the severity of Klein's GLOW syndrome cases and their overgrowth and

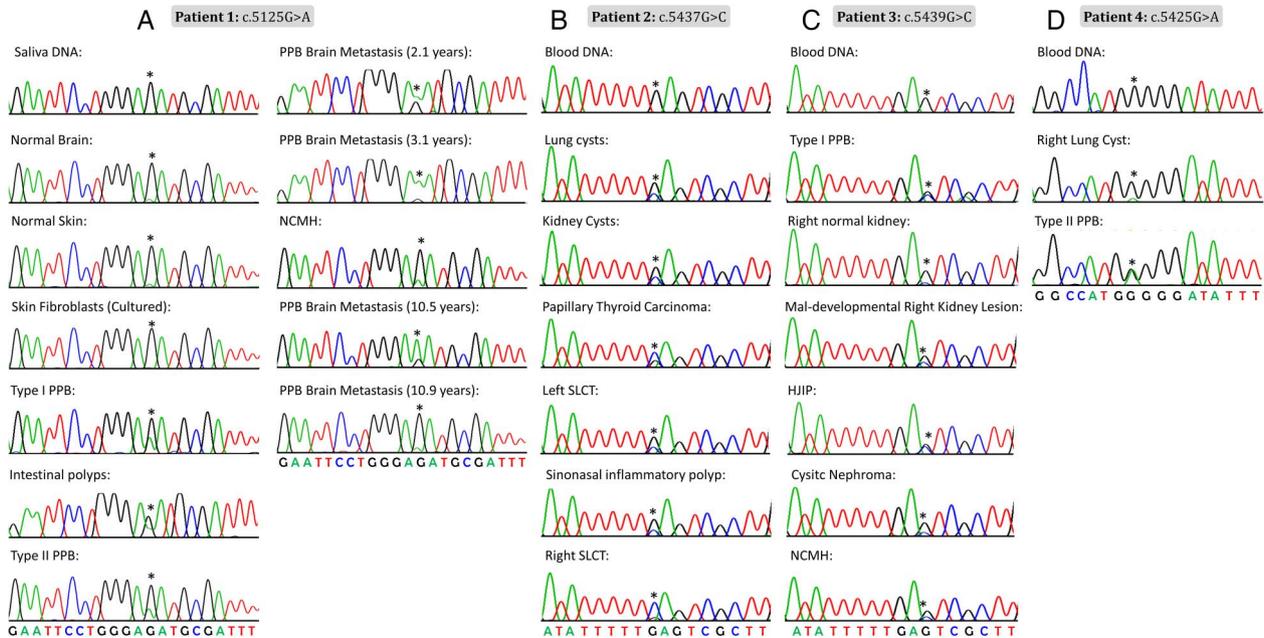


Figure 3 Chromatograms showing the mosaic *DICER1* mutation (indicated by an asterisk) in multiple tissue samples. (A) Patient 1 (c.5125G>A), (B) patient 2 (c.5437G>C), (C) patient 3 (c.5439G>C) and (D) patient 4 (c.5425G>A). HJIP, hamartomatous juvenile intestinal polyps; NCMH, nasal chondromesenchymal hamartoma; PPB, pleuropulmonary blastoma; SLCT, Sertoli–Leydig cell tumour.

developmental delay. Although having several diseases, our cases manifested typical *DICER1* phenotypes, and none had overgrowth or developmental delay. One of Klein's cases had renal dysmorphism, and patient 3 had both CN and contralateral microscopic renal medullary maldevelopment characterised by increased loose mesenchyme, disorganised collecting system and dilated lymphatic vessels, which have not been previously described (see online supplementary data). Both of Klein's patients developed bilateral Wilms tumour; unilateral Wilms and bilateral disease in paired organs are a known feature of *DICER1* syndrome.^{38–41} Rather than comprising a new syndrome, we are inclined to believe that multi-organ mosaic RNase IIIb mutations result in an unusually severe overall *DICER1* phenotype, within which the pleiotropy typical of *DICER1* disease may occasionally result in overgrowth or developmental delay. Identification and analysis of additional mosaic cases may clarify this ambiguity.

Non-RNase IIIb mosaic mutations are likely to exist, and we predict that the phenotype caused by non-RNase IIIb mosaic mutations would be less severe than those caused by mosaic mutations directly affecting the *DICER1* RNase IIIb domain, and therefore, such mutations may be more likely to go undetected. There has been one additional reported instance of a *de novo* germline *DICER1* mutation (c.5125G>C; p. G1709H) affecting a metal-ion binding residue within the RNase IIIb domain.¹² The child is severely affected: he presented at birth with a PitB, extensive multifocal bilateral lung cysts and bilateral renal cystic masses. The c.5125G>C mutation was seemingly heterozygous in lymphocyte gDNA, but extensive investigations to confirm or rule out mosaicism were not possible.¹² The inference from these data is that both mosaic and non-mosaic germline missense mutations affecting exons encoding the metal-ion binding domain of *DICER1* underpin a particularly severe disease phenotype and may induce a large number of disease foci per child, depending on the specific tissue distribution of the mutation (patient 4 might exemplify this more limited yet significant mutation distribution). In

support of this, we recently identified a paternally inherited novel heterozygous germline *DICER1* mutation, c.5441C>T (p. S1814L), in a girl who developed an SLCT and MNG before the age of 13 years (Wu *et al*, unpublished data). This mutation, although located within the RNase IIIb domain, does not directly affect one of the critical catalytic or metal-binding residues within this domain (eg, residues 1705, 1709, 1810 and 1813^{42–43}). The less severe phenotype exhibited by this child may possibly be related to the 'sparing' of the above-mentioned metal-ion binding residues. It is notable that no inherited germline *DICER1* mutations at a nucleotide encoding a metal-ion binding residue have been reported.

Cancer susceptibility syndromes such as familial adenomatous polyposis (FAP) and the neurofibromatoses (NF) are also associated with a mosaic origin of the causative mutations. In these conditions, the disease course of the mosaic form is reported to be milder than the inherited, non-mosaic presentation.^{44–46} For children with mosaic *DICER1* mutations affecting the RNase IIIb domain, the disease appears to be more severe (including earlier onset, greater number of disease foci and greater range of phenotypes) than in the more typical autosomal-dominant forms. This disparity may be attributed to the nature of the mutations required to initiate tumorigenesis in *DICER1* syndrome—typically a first-hit truncating germline mutation occurs in any protein-encoding region and a second 'hit' specifically affects the RNase IIIb domain. Such combinations are likely to be rare since it appears the selected second hit nearly always affects a very limited number of nucleotides encoding the RNase IIIb metal ion-binding domains. In contrast, in the *DICER1* mosaicism reported here, the initial 'hit' is the acquisition of a missense RNase IIIb hotspot mutation. The second likely truncating mutation occurs anywhere across the gene (see online supplementary table S1), and is therefore, stochastically more likely to occur than a RNase IIIb mutation. Thus, we postulate that the combination of the specific effects of the RNase IIIb mutation and widespread inactivating second hits accounts for more severe clinical manifestations in these children.

Table 2 Number of reads containing mutant versus wildtype base at position of interest

Tissue type	Fluidigm Access Array run 1						Fluidigm Access Array run 2						HaloPlex (standard)					HaloPlex ^{HS}					% Average reads							
	Total reads	T	G	C	A	WT%	Mut %	Total reads	T	G	C	A	WT%	Mut %	Total reads	T	G	C	A	WT%	Mut %	Total reads*	T	G	C	A	WT%	Mut %	WT average	Mutant average
Patient 1: c.5125G>A																														
PPB type II†	1102	0	<i>432</i>	0	<u>670</u>	39.20	60.80	6563	4	<i>4404</i>	2	<u>2153</u>	67.10	32.81	10	0	<i>6</i>	0	<u>4</u>	60.0	40.0	73*	0	<i>40</i>	1	<u>32</u>	54.79	43.84	55.27	44.36
Skin fibroblasts (cultured)	5313	3	<i>5027</i>	2	<u>281</u>	94.62	5.29	22 907	10	<i>21 154</i>	4	<u>1739</u>	92.35	7.59	317	0	<i>292</i>	0	<u>24</u>	92.11	7.57	568*	1	<i>541</i>	2	<u>24</u>	95.25	4.23	93.58	6.17
Saliva DNA	5463	5	<i>5178</i>	1	<u>279</u>	94.78	5.11	26 511	6	<i>26 487</i>	4	<u>14</u>	99.91	0.05	1208	0	<i>1184</i>	0	<u>24</u>	98.01	1.99	792*	0	<i>770</i>	0	<u>22</u>	97.22	2.78	97.48	2.48
Patient 2: c.5437G>C																														
Blood DNA	7987	4	<i>7978</i>	0	5	99.89	0	68 572	24	<i>68 503</i>	14	31	99.90	0.02	3223	6	<i>3213</i>	4	0	99.69	0.12	9650*	4	<i>9642</i>	4	0	99.92	0.04	99.85	0.05
Saliva DNA	ND	–	–	–	–	–	–	31 414	6	<i>31 254</i>	139	15	99.49	0.44	795	0	<i>793</i>	1	1	99.75	0.13	1972*	1	<i>1966</i>	5	0	99.70	0.25	99.64	0.27
Urine DNA	ND	–	–	–	–	–	–	14 662	5	<i>14 458</i>	195	4	98.61	1.33	821	0	<i>818</i>	2	1	99.63	0.24	3014*	2	<i>3000</i>	12	0	99.54	0.40	99.26	0.66
Patient 3: c.5439G>C																														
Normal right kidney (#1)†	ND	–	–	–	–	–	–	656	0	<i>656</i>	0	0	100	0	100	0	<i>89</i>	11	0	89.00	11.00	2303*	0	<i>2228</i>	75	0	96.74	3.26	95.25	4.75
Normal right kidney (#2)†	1494	0	<i>1020</i>	474	0	68.27	31.73	3863	1	<i>3791</i>	71	0	98.14	1.84	102	0	<i>90</i>	12	0	88.24	11.76	739*	0	<i>643</i>	95	0	87.01	12.86	85.41	14.55
Blood DNA	ND	–	–	–	–	–	–	ND	–	–	–	–	–	–	ND	–	–	–	–	–	–	2455*	1	<i>2454</i>	0	0	100	0	–	–
Patient 4: c.5425G>A																														
Reactive right lung†	3580	0	<i>3336</i>	0	<u>244</u>	93.18	6.82	7543	0	<i>6541</i>	0	<u>1002</u>	86.72	13.28	1964	2	<i>1937</i>	0	<u>25</u>	98.63	1.27	4401*	1	<i>4091</i>	4	<u>305</u>	92.96	6.93	92.87	7.08
Blood DNA	8266	6	<i>8253</i>	1	<u>6</u>	99.84	0.07	15 501	9	<i>15 481</i>	0	11	99.87	0.00	840	1	<i>838</i>	1	0	99.76	0	16 081*	3	<i>16 073</i>	3	1	99.95	0.01	99.86	0.02
Saliva DNA	ND	–	–	–	–	–	–	ND	–	–	–	–	–	–	647	3	<i>644</i>	0	0	99.54	0	2533*	0	<i>2532</i>	1	0	99.96	0	99.75	0

*Duplicate reads removed (HaloPlex^{HS} data); –, no data. Wildtype allele at position of interest is italicised and the mutant allele is underlined.

†Formalin-fixed paraffin-embedded samples.

Mut, mutant; ND, not done; PPB, pleuropulmonary blastoma, WT, wildtype.

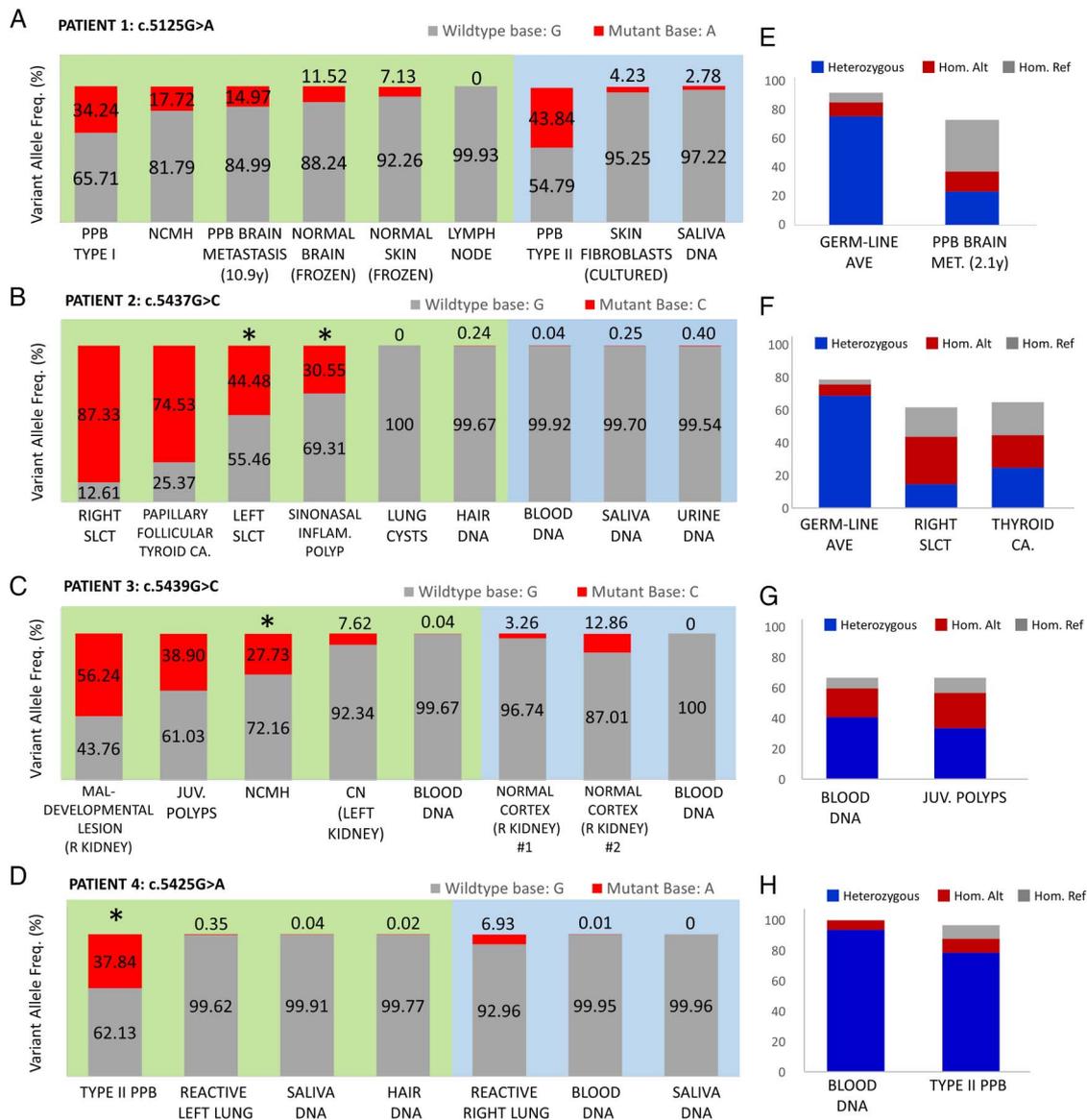


Figure 4 Mutant allele frequencies, second somatic *DICER1* mutations and loss of heterozygosity (LOH). Bar graph indicating the mutant allele fraction detected in multiple tissues from (A) patient 1, (B) patient 2, (C) patient 3 and (D) patient 4. The percentage mutant base fraction (red fill) to wildtype base fraction (grey fill) at the position of interest is indicated for each sample. Green background shading (Left) indicates samples processed using the HaloPlex Standard or Fluidigm Access Array, and blue shading (Right) indicates samples processed using the HaloPlex^{HS} technology. The percentages given for samples sequenced using the HaloPlex Standard or Fluidigm Access Array are averages of all successful runs. Samples designated with an asterisk are tumours found to carry second somatic, likely truncating *DICER1* mutations (see online supplementary table S1 for details). (E–H) Bar graphs illustrating evidence of LOH in tumour samples from patient 1 (E) and patient 2 (F) and lack of LOH in a representative tumour from patient 3 (G) and patient 4 (H). In (E) and (F), there is a notable increase in the number of heterozygous to homozygous SNPs in the tumour samples relative to the germline, which is indicative of LOH. This shift is not evident in tumours from patient 3 (G) and patient 4 (H). CN, cystic nephroma; NCMH, nasal chondromesenchymal hamartoma; PPB, pleuropulmonary blastoma; SLCT, Sertoli–Leydig cell tumour.

Therefore, we predict that non-RNase IIIb mosaic *DICER1* mutations, like mosaic mutations in FAP and NF, will cause a disease phenotype that is milder than both the autosomal-dominant form and that caused by RNase IIIb mosaic *DICER1* mutations.

The importance of identifying the causative mutations in these children is several fold: unaffected parents who have an affected child may want to know the risk of recurrence in future pregnancies. Furthermore, the affected children themselves may want to know the probability of transmission to future offspring. Understanding the genetic cause and the mechanism underlying the phenotype provides information that can be used to ascertain

such risk. Even in heterozygous germline *DICER1* mutation carriers, screening for *DICER1*-related conditions is problematic, as discussed elsewhere.⁶ Mosaicism further complicates such considerations because the distribution of a somatic mutation would be difficult to determine, but it should be borne in mind that these children may be at increased risk compared with other *DICER1* mutation-positive children. In the cases reported here, transmission of the RNase IIIb mutations to the next generation seems unlikely, although in the future, testing of sperm or ova or fetal genetic testing might be considered.

Several high-sensitivity sequencing methods are currently being applied to discover low-frequency mutations, each with its

own advantages and disadvantages. The most prominent so far are the PCR-based Safe-SeqS⁴⁷ and the Molecular Inversion Probe-based smMIP method.⁴⁸ Whole-genome sequencing with higher than average depth of coverage has also been used to identify *de novo* mutations that occurred post-zygotically, as was recently reported by Acuna-Hidalgo *et al.*⁴⁹ Despite these technical advances, detecting low-level mosaicism is still challenging. Low-level mosaic mutations fall below the threshold of sensitivity for many sequencing methods, and other more sensitive technologies are costly and, therefore, may not be practical in either the research or clinical setting. In our hands, the novel HaloPlex^{HS} target enrichment system containing molecular barcodes provided the sensitivity required for detection of mutant allele fractions as low as 0.24%. We found the HaloPlex^{HS} to be an economically feasible platform. It is suitable for covering entire genomic regions (in our case, 499 kb), but currently 5 Mbp is possible, which is in contrast to PCR product and Molecular Inversion Probe-based methods. Additional advantages of HaloPlex^{HS} over the smMIP method include a much lower DNA input requirement (~50 ng) and the redundancy in the HaloPlex^{HS} probe design allows for the vast majority of targeted bases to be covered by at least four probes, ensuring high coverage, but without increasing the cost of the capture. HaloPlex^{HS} is likely to be broadly applicable to other situations where mosaicism can occur but yet remain undetected by currently available technologies. The design implemented can be easily adopted by other investigators interested in identifying mutations in *DICER1* and other genes encoding the components of the miRNA processing machinery. We also demonstrate the utility of the HaloPlex^{HS} in FFPE-derived DNA. Our findings suggest that the targeted ultra-deep next-generation sequencing of the *DICER1* locus is a useful technique for the identification of mosaic *DICER1* mutations.

In summary, by using a new high-sensitivity mutation detection system, we demonstrate that mosaic *DICER1* RNase IIIb missense mutations are an occasional and important genetic cause of the *DICER1* syndrome in patients presenting with multiple primary tumours associated with the syndrome, but for tumour initiation, they often appear to be accompanied by second somatic truncating non-RNase IIIb *DICER1* mutations or LOH.

Author affiliations

¹Department of Human Genetics, McGill University, Montréal, Québec, Canada

²Lady Davis Institute, Segal Cancer Centre, Jewish General Hospital, Montréal, Québec, Canada

³Department of Human Genetics, McGill University and Genome Quebec Innovation Centre, McGill University, Montréal, Québec, Canada

⁴Department of Medical Genetics, Research Institute of the McGill University Health Centre, Montréal, Québec, Canada

⁵Department of Pediatrics, New York Medical College and Maria Fareri Children's Hospital, Valhalla, New York, USA

⁶Department of Pediatric Oncology, Emma Children's Hospital, Academic Medical Center, Amsterdam Zuidoost, The Netherlands

⁷Department of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands

⁸Department of Pediatric Nephrology, Emma Children's Hospital, Academic Medical Center, Amsterdam, The Netherlands

⁹Glick Eye Institute, Indiana University School of Medicine, Indianapolis, Indiana, USA

¹⁰Department of Ophthalmology, University of Louisville, Kentucky, USA

¹¹Agilent Technologies, Santa Clara, California, USA

¹²North Hills, California, USA

¹³Petersburg, Indiana, USA

¹⁴Mahopac, New York, USA

¹⁵Amsterdam, The Netherlands

¹⁶Department of Pathology, CHU-Sainte Justine and University of Montreal, Montréal, Québec, Canada

¹⁷Minneapolis, Minnesota, USA

¹⁸Program in Cancer Genetics, Department of Oncology and Human Genetics, McGill University, Montréal, Québec, Canada

Acknowledgements The authors dedicate this publication to the memory of Sean Callahan. We thank Charmian Cher, Josh Zhiyong Wang and Martin Angers (Agilent Technologies) for their assistance with the HaloPlex^{HS} design and capture and Dr Elizabeth Perlman and Dr Michael McDermott for their assistance with pathology review.

Contributors LdK performed the *DICER1* hotspot sequencing, initial Fluidigm data analysis and mutation validation and wrote the manuscript. LdK, YCW and BR performed the HaloPlex^{HS} capture. TR and DB performed computational analysis. NS performed the initial constitutional gDNA sequencing and MLPA analysis. MW sequenced the tumours of patient 3 and provided some concepts used in the discussion. EW conducted sample acquisition and associated administration work. CS, SMJH, JHMM, JMvH, AHMB, DAP, AR, KLD, TT, JC and CE referred patients, provided samples and collected clinical information. LF created the design of the HaloPlex^{HS}. DB-DS provided expert pathology opinion. JRP reviewed the diagnostic images, collected clinical information and edited the manuscript. JR and WDF designed the study and edited the manuscript. All authors read and approved the final manuscript.

Funding This research was made possible thanks to the support of the Alex's Lemonade Stand Foundation grant to WDF, the Genome Canada Science Technology Innovation Centre, the Compute Canada Resource Allocation Project wst-164-ab and the Genome Innovation Node grants to JR, and the Vanier Canada Graduate Scholarship to LdK.

Competing interests LF is an employee of Agilent Technologies.

Patient consent Depending on the ages of the participants at the time of recruitment, eligible relatives signed a consent form in accordance with the IRB protocol or participants themselves provided written informed consent.

Ethics approval The study was approved by the Institutional Review Board of the Faculty of Medicine of McGill University, Montreal, Quebec, Canada (no. A12-M117-11A).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Sequencing data are available upon request. Please contact the corresponding author.

REFERENCES

- Biesecker LG, Spinner NB. A genomic view of mosaicism and human disease. *Nat Rev Genet* 2013;14:307–20.
- Frank SA. Somatic mosaicism and disease. *Curr Biol* 2014;24:R577–81.
- Lindhurst MJ, Sapp JC, Teer JK, Johnston JJ, Finn EM, Peters K, Turner J, Cannons JL, Bick D, Blakemore L, Blumhorst C, Brockmann K, Calder P, Cherman N, Deardorff MA, Everman DB, Golas G, Greenstein RM, Kato BM, Keppler-Noreuil KM, Kuznetsov SA, Miyamoto RT, Newman K, Ng D, O'Brien K, Rothenberg S, Schwartzenuber DJ, Singhal V, Tirabosco R, Upton J, Wientroub S, Zackai EH, Hoag K, Whitewood-Neal T, Robey PG, Schwartzberg PL, Darling TN, Tosi LL, Mullikin JC, Biesecker LG. A mosaic activating mutation in *AKT1* associated with the Proteus syndrome. *N Engl J Med* 2011;365:611–19.
- Lupski JR. Genetics. Genome mosaicism—one human, multiple genomes. *Science* 2013;341:358–9.
- Hill DA, Ivanovich J, Priest JR, Gurnett CA, Dehner LP, Desruisseau D, Jarzembowski JA, Wikenheiser-Brokamp KA, Suarez BK, Whelan AJ, Williams G, Bracamontes D, Messinger Y, Goodfellow PJ. *DICER1* mutations in familial pleuropulmonary blastoma. *Science* 2009;325:965.
- Foulkes WD, Priest JR, Duchaine TF. *DICER1*: mutations, microRNAs and mechanisms. *Nat Rev Cancer* 2014;14:662–72.
- Klein S, Lee H, Ghahremani S, Kempert P, Ischander M, Teitell MA, Nelson SF, Martinez-Agosto JA. Expanding the phenotype of mutations in *DICER1*: mosaic missense mutations in the RNase IIIb domain of *DICER1* cause GLOW syndrome. *J Med Genet* 2014;51:294–302.
- Doros LA, Turner J, Williams GM, Rossi C, Schultz KAP, Stewart D, Dehner LP, Messinger YH, Hill DA. The *DICER1* syndrome: genotype-phenotype correlation in PPB patients. *J Clin Oncol* 2015. ASCO Annual Meeting (May 29 – June 2, 2015). Vol 33, No 15_suppl (May 20 Supplement), 2015:10022.
- Sabbaghian N, Srivastava A, Hamel N, Plourde F, Gajtko-Metera M, Niedziela M, Foulkes WD. Germ-line deletion in *DICER1* revealed by a novel MLPA assay using synthetic oligonucleotides. *Eur J Hum Genet* 2014;22:564–7.
- Mertes F, Elsharawy A, Sauer S, van Helvoort JM, van der Zaag PJ, Franke A, Nilsson M, Lehrach H, Brookes AJ. Targeted enrichment of genomic DNA regions for next-generation sequencing. *Brief Funct Genomics* 2011;10:374–86.
- Bahubeshi A, Tischkowitz M, Foulkes WD. miRNA processing and human cancer: *DICER1* cuts the mustard. *Sci Transl Med* 2011;3:111ps46.

- 12 de Kock L, Sabbaghian N, Plourde F, Srivastava A, Weber E, Bouron-Dal Soglio D, Hamel N, Choi JH, Park S-H, Deal CL, Kelsey MM, Dishop MK, Esbenshade A, Kuttesch JF, Jacques TS, Perry A, Leichter H, Maeder P, Brundler M-A, Warner J, Neal J, Zacharin M, Korbonits M, Cole T, Traunecker H, McLean TW, Rotondo F, Lepage P, Albrecht S, Horvath E, Kovacs K, Priest JR, Foulkes WD. Pituitary blastoma: a pathognomonic feature of germ-line DICER1 mutations. *Acta Neuropathol* 2014;128:111–22.
- 13 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–20.
- 14 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
- 15 Illumina. igenomes, ready-to-use reference sequences and annotations. http://support.illumina.com/sequencing/sequencing_software/igenome.html
- 16 Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* 2015;31:2032–4.
- 17 Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics* 2011;27:2987–93.
- 18 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–9.
- 19 Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26:841–2.
- 20 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
- 21 DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491–8.
- 22 Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013;11:11.10.11–33.
- 23 Durbin R, Danecek P, Schifffels S. Multiallelic calling model in bcftools (-m). June 2014. <http://samtools.github.io/bcftools/call-m.pdf>
- 24 Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012;6:80–92.
- 25 Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29:308–11.
- 26 Liu X, Jian X, Boerwinkle E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat* 2011;32:894–9.
- 27 Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum Mutat* 2013;34:E2393–2402.
- 28 Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, Lu X. Using *Drosophila melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front Genet* 2012;3:35.
- 29 Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS Comput Biol* 2013;9:e1003153.
- 30 Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol* 2011;29:24–6.
- 31 Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 2013;14:178–92.
- 32 Nur S, Badr R, Sandoval C, Brudnicki A, Yeh A. Syndromic presentation of a pleuropulmonary blastoma associated with congenital cystic adenomatoid malformation. A case report. *J Pediatr Surg* 2007;42:1772–5.
- 33 Ramasubramanian A, Correa ZM, Augsburger JJ, Sisk RA, Plager DA. Medulloepithelioma in DICER1 syndrome treated with resection. *Eye* 2013;27:896–7.
- 34 Anglesio MS, Wang Y, Yang W, Senz J, Wan A, Heravi-Moussavi A, Salamanca C, Maines-Bandiera S, Huntsman DG, Morin GB. Cancer-associated somatic DICER1 hotspot mutations cause defective miRNA processing and reverse-strand expression bias to predominantly mature 3p strands through loss of 5p strand cleavage. *J Pathol* 2013;229:400–9.
- 35 Murray MJ, Bailey S, Raby KL, Saini HK, de Kock L, Burke GA, Foulkes WD, Enright AJ, Coleman N, Tischkowitz M. Serum levels of mature microRNAs in DICER1-mutated pleuropulmonary blastoma. *Oncogenesis* 2014;3:e87.
- 36 Pugh TJ, Yu W, Yang J, Field AL, Ambrogio L, Carter SL, Cibulskis K, Giannikopoulos P, Kiezun A, Kim J, McKenna A, Nickerson E, Getz G, Hoffner S, Messinger YH, Dehner LP, Roberts CWM, Rodriguez-Galindo C, Williams GM, Rossi CT, Meyerson M, Hill DA. Exome sequencing of pleuropulmonary blastoma reveals frequent biallelic loss of TP53 and two hits in DICER1 resulting in retention of 5p-derived miRNA hairpin loop sequences. *Oncogene* 2014;33:5295–302.
- 37 Behjati S, Maschietto M, Williams RD, Side L, Hubank M, West R, Pearson K, Sebire N, Tarpey P, Futreal A, Brooks T, Stratton MR, Anderson J. A pathogenic mosaic TP53 mutation in two germ layers detected by next generation sequencing. *PLoS ONE* 2014;9:e96531.
- 38 Foulkes WD, Bahubeshi A, Hamel N, Pasini B, Asioli S, Baynam G, Choong CS, Charles A, Frieder RP, Dishop MK, Graf N, Ekim M, Bouron-Dal Soglio D, Arseneau J, Young RH, Sabbaghian N, Srivastava A, Tischkowitz MD, Priest JR. Extending the phenotypes associated with DICER1 mutations. *Hum Mutat* 2011;32:1381–4.
- 39 Torrezan GT, Ferreira EN, Nakahata AM, Barros BD, Castro MT, Correa BR, Krepischi AC, Olivieri EH, Cunha IW, Tabori U, Grundy PE, Costa CM, de Camargo B, Galante PA, Carraro DM. Recurrent somatic mutation in DROSHA induces microRNA profile changes in Wilms tumour. *Nat Commun* 2014;5:4039.
- 40 Bahubeshi A, Bal N, Rio Frio T, Hamel N, Pouchet C, Yilmaz A, Bouron-Dal Soglio D, Williams GM, Tischkowitz M, Priest JR, Foulkes WD. Germline DICER1 mutations and familial cystic nephroma. *J Med Genet* 2010;47:863–6.
- 41 Wu MK, Sabbaghian N, Xu B, Addidou-Kalucki S, Bernard C, Zou D, Reeve AE, Eccles MR, Cole C, Choong CS, Charles A, Tan TY, Iglesias DM, Goodyer PR, Foulkes WD. Biallelic DICER1 mutations occur in Wilms tumours. *J Pathol* 2013;230:154–64.
- 42 Takeshita D, Zenno S, Lee WC, Nagata K, Saigo K, Tanokura M. Homodimeric structure and double-stranded RNA cleavage activity of the C-terminal RNase III domain of human dicer. *J Mol Biol* 2007;374:106–20.
- 43 Heravi-Moussavi A, Anglesio MS, Cheng S-W, Senz J, Yang W, Prentice L, Fejes AP, Chow C, Tone A, Kalloger SE, Hamel N, Roth A, Ha G, Wan AN, Maines-Bandiera S, Salamanca C, Pasini B, Clarke BA, Lee AF, Lee C-H, Zhao C, Young RH, Aparicio SA, Sorensen PH, Woo MM, Boyd N, Jones SJ, Hirst M, Marra MA, Gilks B, Shah SP, Foulkes WD, Morin GB, Huntsman DG. Recurrent somatic DICER1 mutations in nonepithelial ovarian cancers. *N Engl J Med* 2012;366:234–42.
- 44 Evans DG. Neurofibromatosis type 2 (NF2): a clinical and molecular review. *Orphanet J Rare Dis* 2009;4:16.
- 45 Hes FJ, Nielsen M, Bik EC, Konvalinka D, Wijnen JT, Bakker E, Vasen HF, Breuning MH, Tops CM. Somatic APC mosaicism: an underestimated cause of polyposis coli. *Gut* 2008;57:71–6.
- 46 Consoli C, Moss C, Green S, Balderson D, Cooper DN, Upadhyaya M. Gonosomal mosaicism for a nonsense mutation (R1947X) in the NF1 gene in segmental neurofibromatosis type 1. *J Invest Dermatol* 2005;125:463–6.
- 47 Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci USA* 2011;108:9530–5.
- 48 Hiatt JB, Pritchard CC, Salipante SJ, O’Roak BJ, Shendure J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. *Genome Res* 2013;23:843–54.
- 49 Acuna-Hidalgo R, Bo T, Kwint MP, van de Vorst M, Pinelli M, Veltman JA, Hoischen A, Visser LE, Gilissen C. Post-zygotic point mutations are an underrecognized source of de novo genomic variation. *Am J Hum Genet* 2015;97:67–74.

SUPPLEMENTAL NOTE: CASE REPORTS

Patient 1:

The early clinical presentation of patient 1 is described by Nur *et al* (2007)¹. Bilateral lung cysts were detected at 35 weeks of gestation. The infant underwent surgery at 5 days of age to remove the cystic lung lesions, which although initially classified as a congenital cystic adenomatoid malformation (CCAM) type 4, were later revised to a Type I PPB. At 9 months of age, the boy presented with bilious vomiting and abdominal pain. Ileocolic intussusception was detected to have been caused by multiple small bowel polyps consistent with enteritis cystica profunda, which were subsequently removed at 1.7 years. At 25 months of age, a type II PPB was detected to have arisen in a pre-existing cystic lung lesion. The infant underwent a right thoracotomy. The patient also developed four independent metastatic PPB lesions in the brain at ages 2.1 years, 3.1 years, 10.5 years and 10.9 years respectively. The brain tumours were surgically resected and chemotherapy and radiation were implemented. At 8 years of age the boy developed a nasal chondromesenchymal hamartoma (NCMH) which was surgically excised. Very recently, at the age of 12.5 years, the boy was noted to have another recurrence of the brain tumour.

Patient 2:

The clinical presentation of patient 2 was first reported by Ramasunramanian *et al* in 2013². This female child developed bilateral renal cysts which were first noted at age 13 months due to abdominal distension. The left kidney was biopsied at 1 year 2 months of age and upon pathologic review, the lesions were diagnosed as hamartomatous renal cysts, not otherwise specified. Bilateral, multifocal lung cysts were noted incidentally on abdominal X-Ray for renal abnormalities and at the age of 2.5 years, the lung cysts were surgically removed. A deformity of the left eye, described as a “pre-phthisical” eye, was present at birth. At approximately 2 years of age, the girl underwent surgery to correct the deformity and remove a vascular mass and the retinal detachment was repaired. Several years later, unreparable retinal detachment occurred in the left eye. At the age of 7.7 years, morning symptoms of nausea and vomiting were noted. A tumour of the pineal gland (pineoblastoma) was

identified and the child underwent a 6 month course of chemotherapy and radiation following surgical resection. A thyroid tumour was detected at 10.6 years of age. The patient underwent a thyroidectomy and the tumour was diagnosed as a follicular variant papillary thyroid carcinoma. At 13.4 years of age, a left ovarian tumour was noted incidentally on renal follow-up. The tumour was resected and diagnosed as a Sertoli-Leydig cell tumour (SLCT). An unusual hamartomatous nasal polyp, not otherwise characterised, developed at 15.1 years of age which was surgically removed. A right ovarian SLCT was detected and surgically excised at age 15.7 years and at the age of 17.2 years, the girl developed a teratoid ciliary body medulloepithelioma (CBME) (previously described²) which was surgically removed. The patient is currently 20.5 years of age and is disease free.

Patient 3:

The female infant underwent surgical resection of right pulmonary cystic lesions, diagnosed as CCAM (later revised to Type I PPB), on the 10th day of life. Lung cysts were also detected in the left lung at the time. At 1 month of age, a right renal medullary malformation with disorganized collecting system and dilated lymphatic vessels was detected and at 6 months of age, the patient developed hamartomatous juvenile intestinal polyps. The young girl then developed a cystic nephroma of the left kidney at 1.1 years of age and at 6.8 years of age, a nasal chondromesenchymal hamartoma was detected and removed. At the time of writing, the girl is 9.4 years old and is disease free.

Patient 4:

At 11 months of age, the male infant was found to have bilateral multifocal lung cysts. Following surgical resection, the lesions were diagnosed as CCAM, but were later revised to Type I PPB. At 1.7 years of age, he presented with a right infected lung cyst and at 4.1 years of age, a left residual Type I PPB with polypoid Type II PPB was excised. At the time of writing, the boy is 16.7 years old and is disease free.

SUPPLEMENTAL FIGURES

Figure S1:

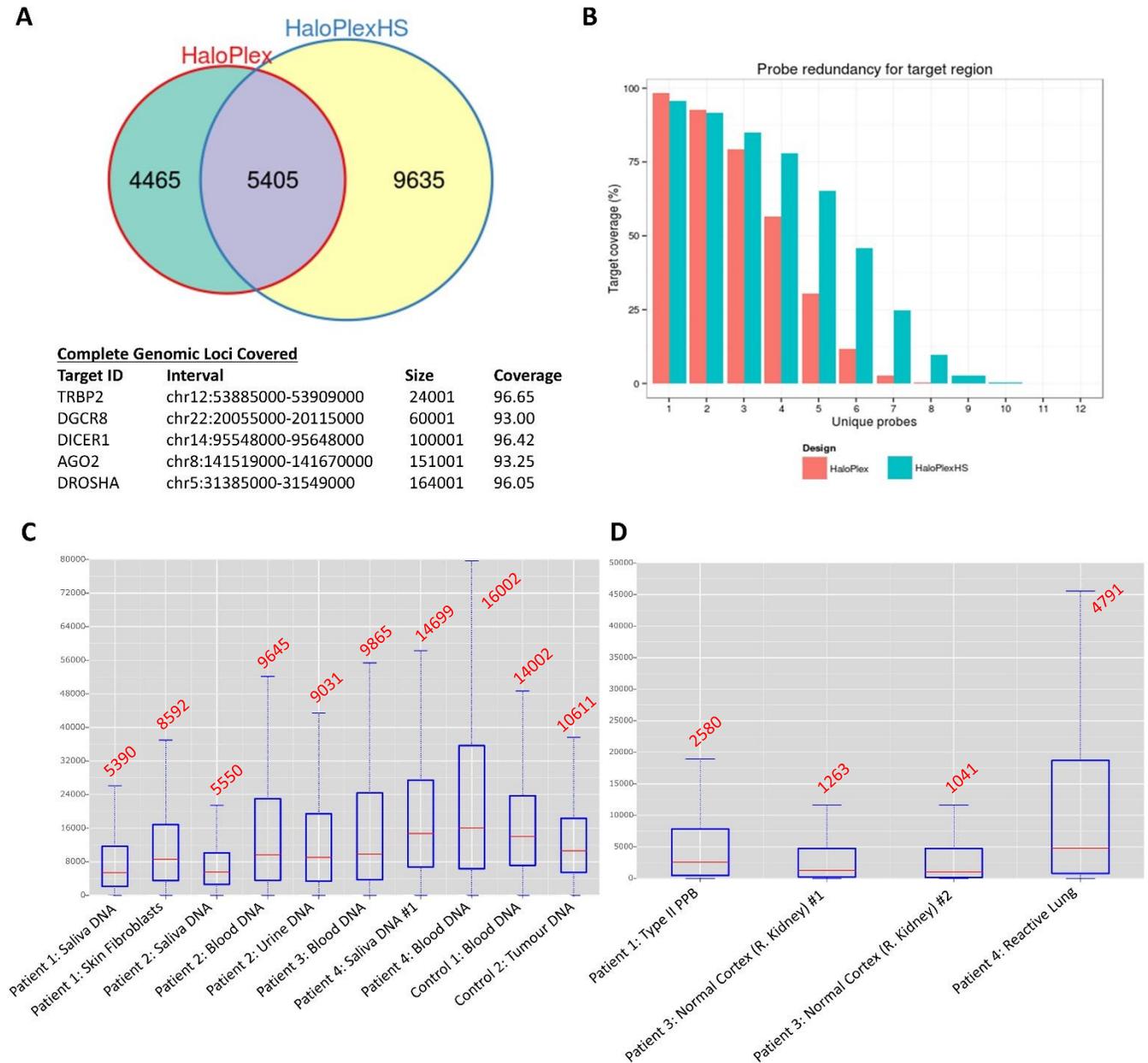


Figure S1 Legend: Assay metrics. **A)** Probe overlap between standard HaloPlex and HaloPlex^{HS} design and details of complete genomic loci covered. **B)** Graph depicting greater amplicon redundancy in HaloPlex^{HS} design compared to the standard HaloPlex. Distribution of reads obtained for all non-formalin-fixed paraffin-embedded (FFPE) samples **(C)** and FFPE samples **(D)** obtained following capture with the HaloPlex^{HS}. The median number of reads per base is indicated by the red bar and is presented in red text above each plot.

Figure S2:

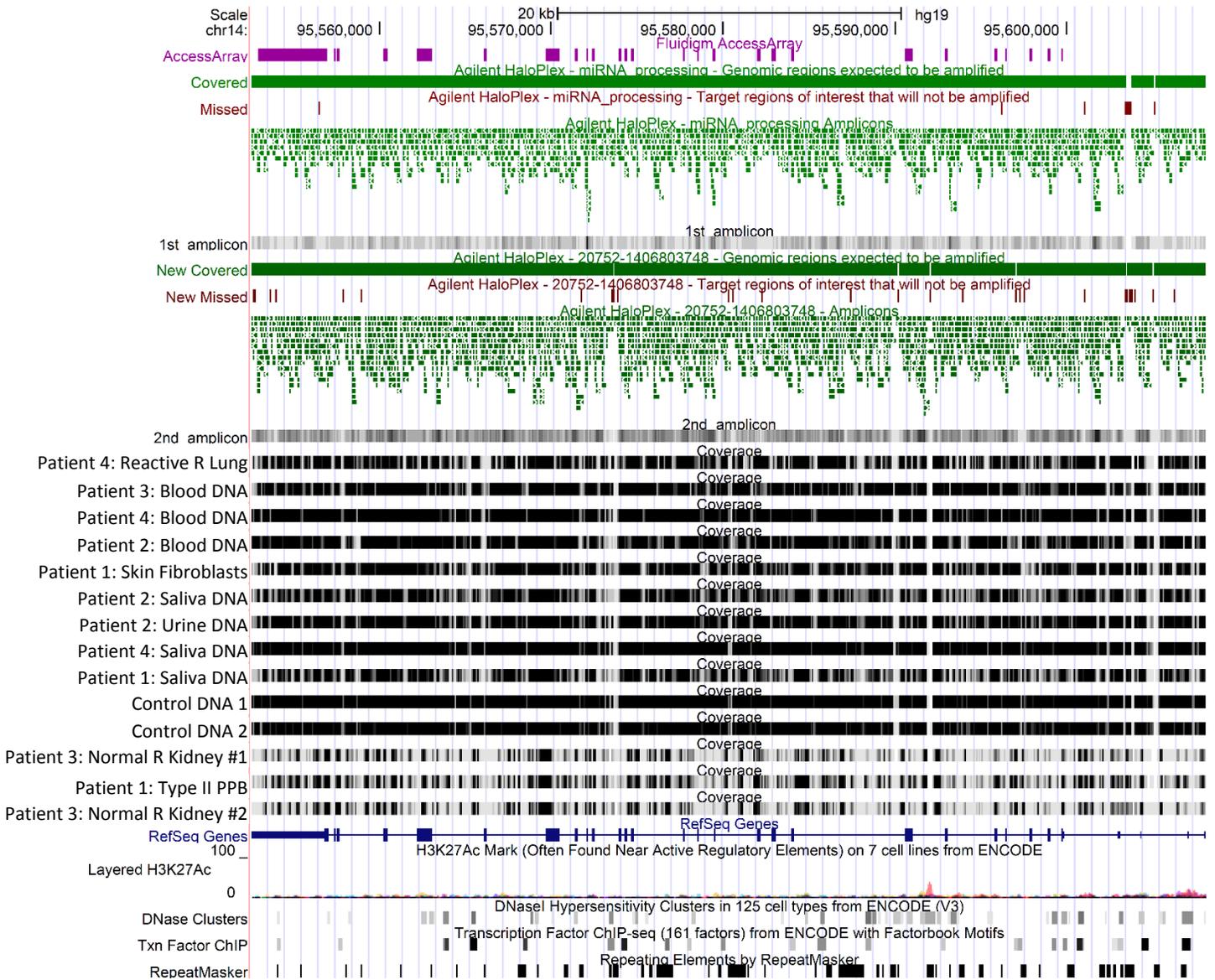


Figure S2 Legend: UCSC Genome Browser visualization of the custom Fluidigm Access array, HaloPlex standard and HaloPlex^{HS} capture designs, focussing on the *DICER1* locus. From top to bottom, target regions of the Fluidigm Access array are shown in purple; target regions include all exons and exon-intron boundaries of *DICER1*. The following four tracks show the coverage of the *DICER1* target region (green) for the HaloPlex standard capture, the regions not ultimately covered (maroon), the HaloPlex standard amplicon distribution

(green) and the degree of design redundancy in greyscale (1st amplicon). The same is shown for the HaloPlex^{HS} capture below: The regions not ultimately covered are shown in maroon and the HaloPlex^{HS} amplicon distribution in green. Note the greater amplicon redundancy in the HaloPlex^{HS} design as expressed in the grey bar (2nd amplicon). The subsequent thirteen grey tracks depict the overall coverage obtained for each of the samples following capture with the HaloPlex^{HS}. The position of the *DICER1* exons (Refseq genes) is shown in blue, followed by regulatory data: H3K27Ac, DNase I hypersensitivity clusters and transcription factor binding sites from ENCODE data, as well as the position of repeat elements (repeat masker). The regulatory elements are distributed over the introns of the gene indicating that SNPs or mutations in these regions may have regulatory effects.

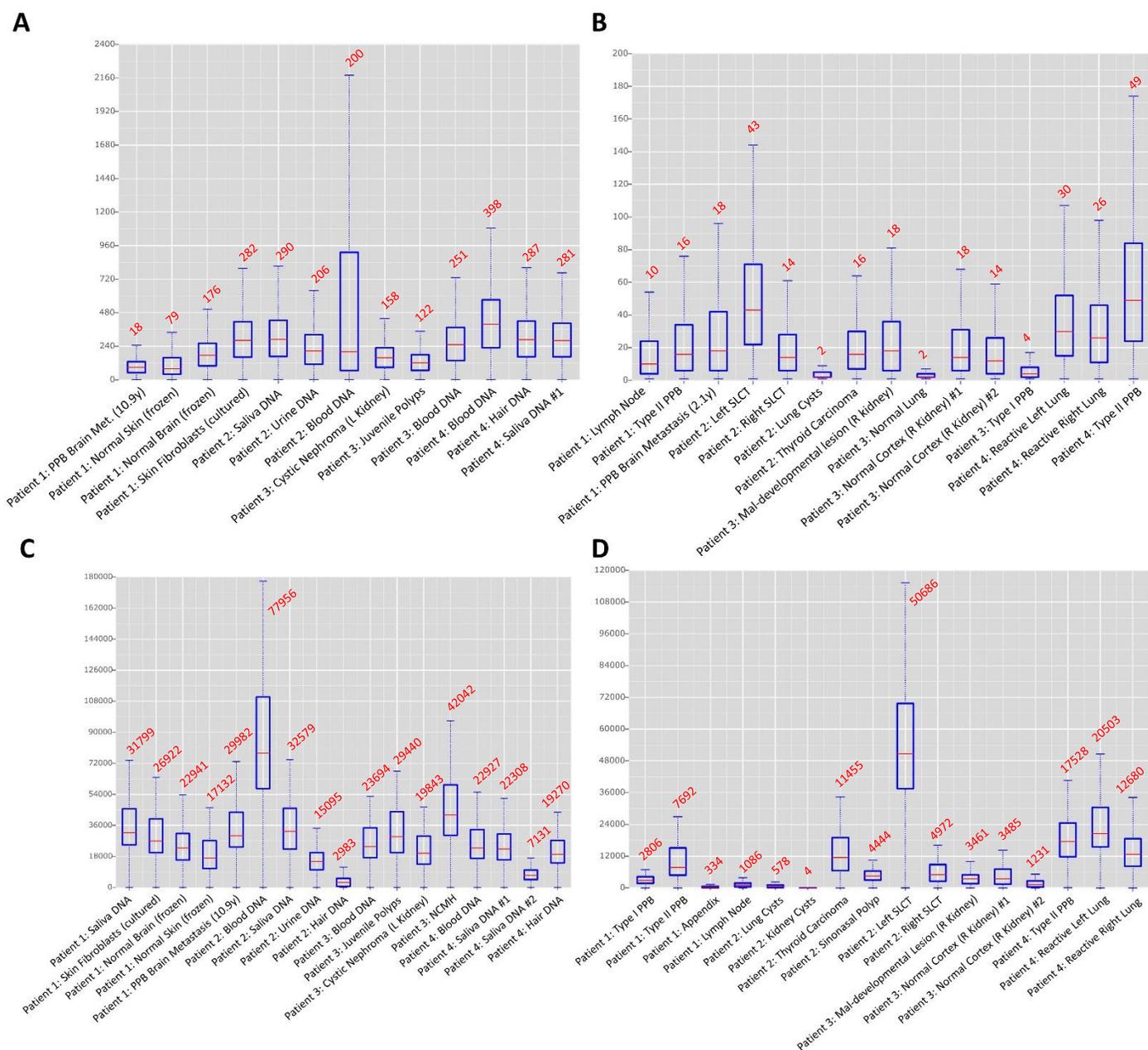


Figure S3:

Figure S3 Legend: Box plot of distribution of coverage achieved per base for the HaloPlex standard and Fluidigm Access Array-derived datasets. Distribution of coverage achieved per base for all non-formalin-fixed paraffin-embedded (FFPE) samples (A) and FFPE samples (B) obtained following capture with the HaloPlex standard shown in reads per sample. Depth of coverage of all non-FFPE samples (C) and FFPE samples

(D) obtained following capture with the Fluidigm Access Array is shown in reads per sample. The depth shown in **C** and **D** is the sum of two successive runs. The median number of reads per base is indicated by the red bar and is presented in red text above each plot.

Figure S4:

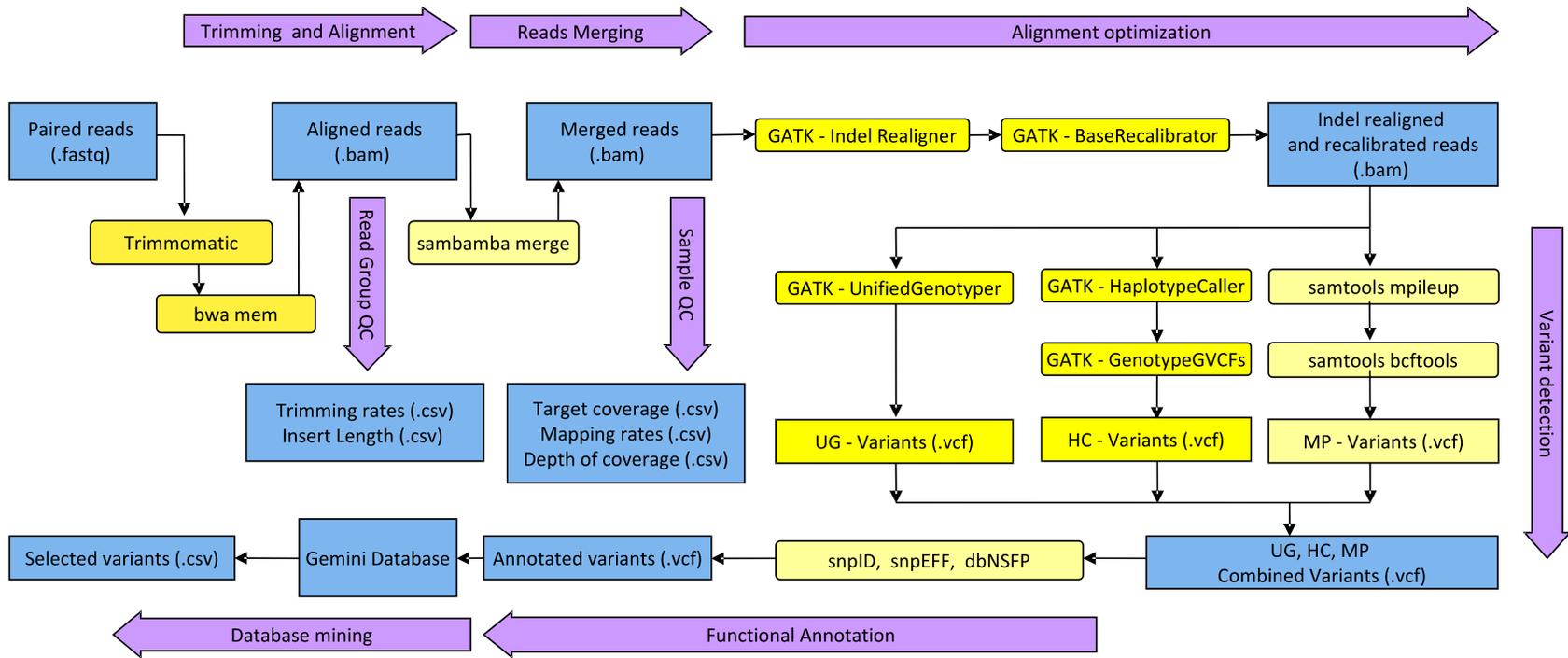
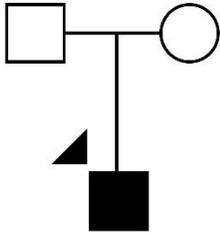


Figure S4 Legend: Overview of bioinformatics methodologies and custom pipeline for genome alignment, variant detection, functional annotation and selection. Arrows represent workflow. Rounded rectangles represent data processing using various software packages. Square boxes represent output files from the pipeline

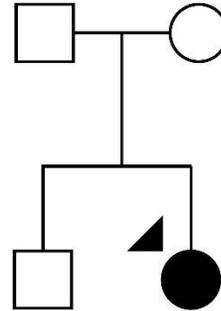
Figure S5:

A



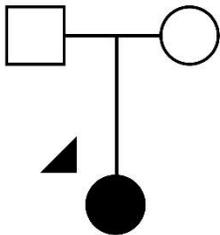
- *in utero Bilateral lung cysts
- *birth; 5 days Bilateral CCAM, revised later to Type I PPB
- *birth Scrotal web and testicular cyst
- *9 m; 1.7 y Multiple small bowel polyps: enteritis cystica profunda
- *2 y Type II PPB
- *2 y PPB brain metastasis #1
- 3.1 y PPB brain metastasis #2
- 8 y NCMH
- 10.5 y PPB brain metastasis #3
- 10.9 y PPB brain metastasis #4

B



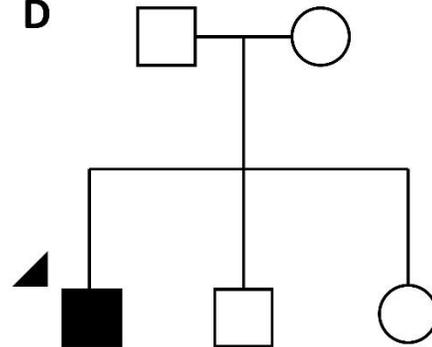
- *13 m; 1.2 y Hamartomatous bilateral renal cysts, NOS
- *13 m; 2.5 y Benign multifocal bilateral lung cysts, NOS
- *birth; 2 y L ocular 'pre-phthisical changes': vascular mass + recurrent retinal detachments
- 7.7 y Pineoblastoma
- 10.6 y Follicular variant of papillary thyroid tumor
- 13.4 y L ovarian SLCT
- 15.1 y Unusual hamartomatous nasal polyp, not otherwise categorized
- 15.7 y R ovarian SLCT
- *16 y; 17.2 y R teratoid CBME

C



- 10 d R CCAM, revised to Type I PPB; L lung cysts
- 6 m HJIP
- 1 m R renal medullary malformation with disorganized collecting system and dilated lymphatic vessels
- 1.1 y L CN
- 6.8 y NCMH

D



- *11 m bilateral multifocal lung cysts; L CCAM; revised to Type I PPB
- 1.7 y R infected pulmonary cyst
- 4.1 y L residual Type I PPB with polypoid Type II PPB

Figure S5: Family pedigrees. Pedigree of patient 1 (A), patient 2 (B) and patient 3 (C) and patient 4 (D) showing unremarkable family histories.

Figure S6:

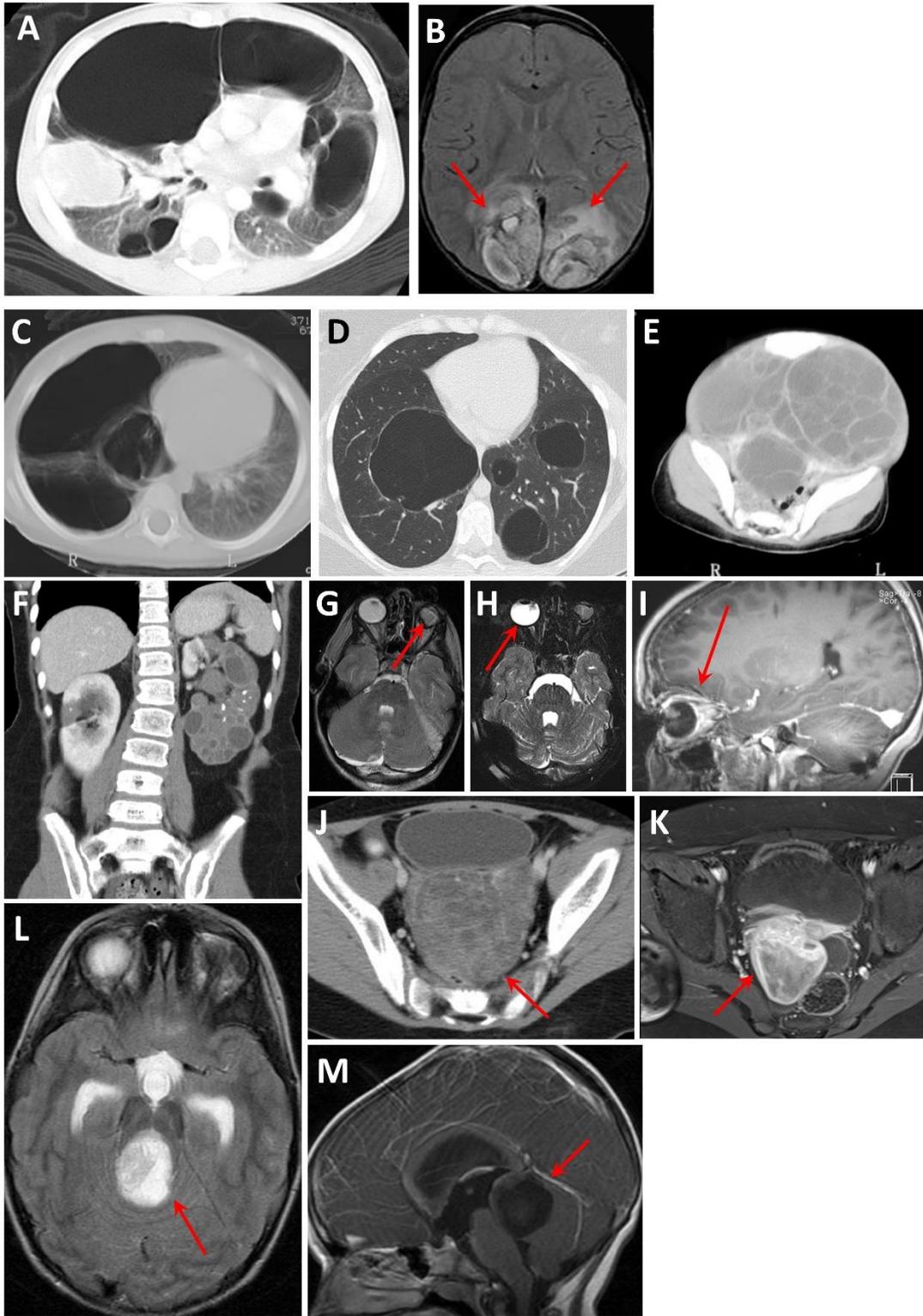


Figure S6 Legend: Diagnostic images from Patient 1 (A and B) and Patient 2 (C to M). **A)** An axial computed tomography (CT) image of the type II PPB that arose in a type I air-filled cyst. **B)** A CT image showing the bi-occipital metastatic PPB lesion (indicated by arrows) at age 2.1 years in Patient 1. **C)** Axial computed tomography (CT) image of the pulmonary lesions (likely PPB Type I and Type Ir) at age 13 months. Note cardiac shift. **D)** Pulmonary Lesions at age 19 years: axial CT images showing residual Type Ir cysts. **E)** Massive bilateral cystic renal lesions at age 24 months. Note massive protrusion at the level of the pelvis. **F)** Persisting renal cysts seen at age 12 years. **G)** Deformity of left eye at age 11 years and **H and I)** ciliary body medulloepithelioma (CBME) of the left eye at 17.2 years of age. **J)** Ovarian Sertoli-Leydig cell tumour of the left ovary (age 13.4 years) and **K)** of the right ovary (age 15.7 years). **L and M)** Posterior fossa PNET (pineoblastoma) at age 7.7 years.

Figure S7:

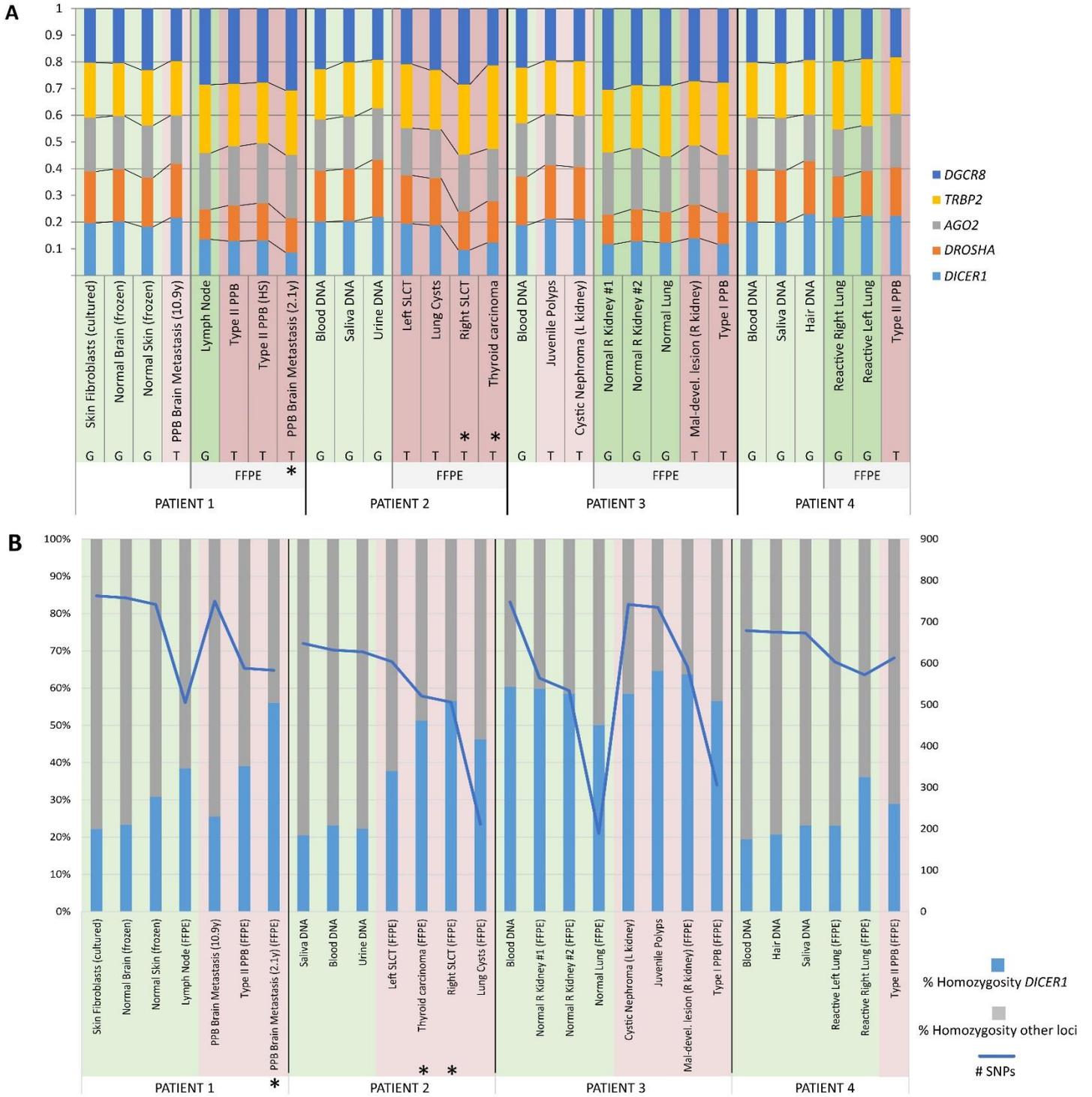


Figure S7 Legend: Measure of potential copy number loss or loss of heterozygosity. **A)** Comparison of the average coverage achieved for all loci. The relative values of the average coverage for each loci are presented in the graph. A perfectly homogenous coverage across all loci would be shown as increments of 0.2. The samples are separated firstly by patient and subsequently by whether they are formalin-fixed paraffin embedded (FFPE) or not, as the FFPE treatment introduces a bias towards loci with smaller amplicons. Asterisks indicate samples that show evidence of either copy number loss or significant loss of heterozygosity. **B)** Comparison of the homozygosity of the *DICER1* locus to the other targeted loci. The bar graph represents the number of homozygous SNPs within *DICER1* relative to total SNPs and proportionally to the number of homozygous SNPs of the other loci (see table S2 for precise numbers). The line graph indicates the total number of SNPs found per sample, which gives an indication of the sequencing depth/quality of a particular sample. Asterisks indicate samples that show evidence of either copy number loss or significant loss of heterozygosity. Despite there being an apparent increase in the percent homozygosity of SNPs in *DICER1* in the lung cyst of Patient 2 (graph B, Patient 2, column 7), there was no visible decrease in the average coverage of the *DICER1* locus in the tumor (graph A, Patient 2, column 5), and we therefore considered there to be no LOH in Patient 2's lung cyst. *Abbreviations:* FFPE, formalin-fixed paraffin embedded; G, germ-line; PPB, pleuropulmonary blastoma; R, right; SLCT, Sertoli-Leydig cell tumor; T, tumor.

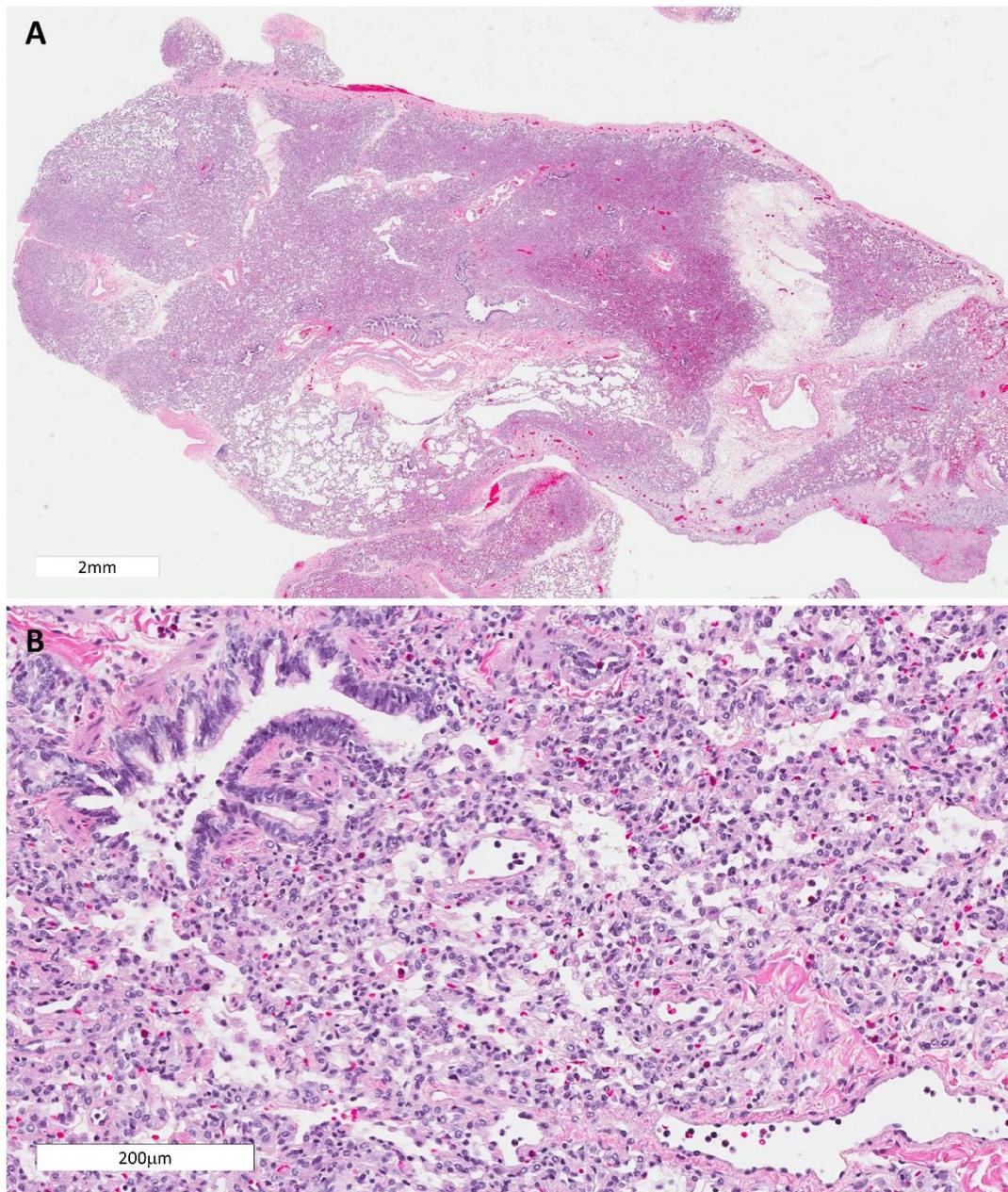


Figure S8 Legend: Haematoxylin and eosin stained images of Patient 4's reactive lung lesion. A) Image at 10x magnification showing congestive lung tissue. Areas that appear dense are due to lack of inflation of the lung tissue. No tumour cells are visible within these regions. **B)** A 200x magnified view of the reactive lung tissue. A bronchiole is present at the top left of the image, numerous macrophages are seen in the alveoli and no tumour cells are visible.

Table S1: number of reads containing mutant vs wild-type base at position of interest and second “hit” data

Tissue Type	Fluidigm Access Array 1							Fluidigm Access Array 2							HaloPlex (Standard)						% Average Reads		Second "Hit" Data				
	Total Reads	T	G	C	<u>A</u>	WT %	Mut %	Total Reads	T	G	C	<u>A</u>	WT %	Mut %	Total Reads	T	G	C	<u>A</u>	WT %	Mut %	WT ave.	Mutant ave.	DNA Change	Protein Change		
PATIENT 1: c.5125G>A																											
PPB Type I ^a	2342	1	<i>1539</i>	0	<u>802</u>	65.71	34.24	Failed	-	-	-	-	-	-	ND	-	-	-	-	-	-	-	-	-	NI	NI	
Lymph Node ^a	1461	0	<i>1460</i>	1	<u>0</u>	99.93	0	Failed	-	-	-	-	-	-	ND	-	-	-	-	-	-	-	-	-	NA	NA	
NCMH ^a	ND	-	-	-	-	-	-	1219	6	<i>997</i>	0	<u>216</u>	81.79	17.72	ND	-	-	-	-	-	-	-	-	-	NI	NI	
PPB Brain metastasis (10.9y)	1756	0	<i>1517</i>	1	<u>238</u>	86.39	13.55	27901	13	<i>22697</i>	3	<u>5188</u>	81.35	18.59	94	0	<i>82</i>	0	<u>12</u>	87.23	12.77	84.99	14.97	LOH			
Normal brain (frozen)	4023	5	<i>3630</i>	0	<u>388</u>	90.23	9.64	19243	11	<i>16616</i>	2	<u>2614</u>	86.35	13.58	194	1	<i>171</i>	0	<u>22</u>	88.14	11.34	88.24	11.52	NA	NA		
Normal skin (frozen)	1992	6	<i>1986</i>	0	<u>0</u>	99.7	0	11988	4	<i>10493</i>	1	<u>1490</u>	87.53	12.43	67	0	<i>60</i>	0	<u>6</u>	89.55	8.96	92.26	7.13	NA	NA		
PATIENT 2: c.5437G>C																											
Lung cysts ^a	822	0	<i>822</i>	<u>0</u>	0	100	0	Failed	-	-	-	-	-	-	ND	-	-	-	-	-	-	-	-	-	-	NI	
Kidney cysts ^a	6	0	<i>4</i>	<u>0</u>	2	66.67	0.00	Failed	-	-	-	-	-	-	ND	-	-	-	-	-	-	-	-	-	-	LOH	
Papillary Follicular Thyroid Ca. ^a	1887	3	<i>741</i>	<u>1143</u>	0	39.27	60.57	14079	6	<i>315</i>	<u>13747</u>	11	2.24	97.64	52	0	<i>18</i>	<u>34</u>	0	34.62	65.38	25.37	74.53	LOH			
Left SLCT ^a	4625	1	<i>2603</i>	<u>2018</u>	3	56.28	43.63	46179	18	<i>25513</i>	<u>20630</u>	18	55.25	44.67	237	0	<i>130</i>	<u>107</u>	0	54.85	45.15	55.46	44.48	c.4626_4626delG	p.Q1542Hfs*18		
Sinonasal Inflammatory Polyp ^a	4445	6	<i>3081</i>	<u>1358</u>	0	69.31	30.55	Failed	-	-	-	-	-	-	ND	-	-	-	-	-	-	-	-	-	c.4458_4458delA	p.K1486Nfs*4	
Right SLCT ^a	1721	1	<i>169</i>	<u>1551</u>	0	9.82	90.12	4214	2	<i>20</i>	<u>4189</u>	3	0.47	99.41	69	0	<i>19</i>	<u>50</u>	0	27.54	72.46	12.61	87.33	LOH			
Hair DNA	ND	-	-	-	-	-	-	5800	3	<i>5781</i>	<u>14</u>	2	99.67	0.24	ND	-	-	-	-	-	-	-	-	-	-	NA	NA
PATIENT 3: c.5439G>C																											
Juvenile Polyps	6385	5	<i>3614</i>	<u>2764</u>	2	56.60	43.29	23155	9	<i>13709</i>	<u>9420</u>	17	59.21	40.68	272	0	<i>183</i>	<u>89</u>	0	67.28	32.72	61.03	38.90	NI	NI		
Cystic Nephroma (L Kidney)	6350	3	<i>5747</i>	<u>598</u>	2	90.50	9.42	12277	6	<i>11465</i>	<u>804</u>	2	93.39	6.55	407	0	<i>379</i>	<u>28</u>	0	93.12	6.88	92.34	7.62	NI	NI		
Mal-developmental R Kidney	ND	-	-	-	-	-	-	2194	0	<i>178</i>	2016	0	8.11	91.89	136	0	<i>108</i>	28	0	79.41	20.59	43.76	56.24	NA	NA		
NCMH	6535	6	<i>4648</i>	<u>1879</u>	2	71.12	28.75	34561	17	<i>25296</i>	<u>9232</u>	16	73.19	26.71	ND	-	-	-	-	-	-	72.16	27.73	c.4651_4652insTGCT	p.E1551Vfs*7		
Blood DNA	6373	3	<i>6361</i>	<u>8</u>	1	99.81	0.13	16088	40	<i>16030</i>	<u>0</u>	18	99.64	0	681	2	<i>678</i>	<u>0</u>	1	100	0	99.67	0.04	NA	NA		
PATIENT 4: c.5425G>A																											
Type II PPB ^a	3972	1	<i>2515</i>	1	<u>1455</u>	63.32	36.63	13160	4	<i>7885</i>	1	<u>5270</u>	59.92	40.05	285	0	<i>180</i>	0	<u>105</u>	63.16	36.84	62.13	37.84	c.1966C>T	p.R656*		
Reactive Left lung ^a	3442	1	<i>3405</i>	0	<u>36</u>	98.93	1.05	14882	7	<i>14871</i>	2	<u>2</u>	99.93	0.01	197	0	<i>197</i>	0	<u>0</u>	100	0	99.62	0.35	NA	NA		
Saliva DNA	6796	6	<i>6788</i>	2	<u>0</u>	99.88	0	19702	6	<i>19685</i>	2	<u>9</u>	99.91	0.05	ND	-	-	-	-	-	-	-	-	-	NA	NA	
Hair DNA	ND	-	-	-	-	-	-	17087	4	<i>17075</i>	0	<u>8</u>	99.93	0.05	765	3	<i>762</i>	0	<u>0</u>	99.61	0	99.77	0.02	NA	NA		

Abbreviations: LOH, loss of heterozygosity; Mut, mutant; NA, not applicable; NCMH, nasal chondromesenchymal hamartoma; ND, not done; NI, none identified; PPB, pleuropulmonary blastoma, SLCT, Sertoli-Leydig cell tumour; WT, wild-type. **Notes:** a Formalin-fixed paraffin-embedded samples; - No data.

Legend: Number of reads containing mutant vs wild-type base at position of interest and second “hit” data. Wild-type allele at the position of interest is *italicised* and the mutant allele is underlined (e.g. Patient 1: wild-type = G; mutant = A). Samples that were sequenced, but did not have sufficient coverage are not presented in the table.

TABLE S1

TABLE S2

Table S2: Total number of SNPs and percentage of homozygosity in *DICER1* and other loci

Patient	Sample	<i>DICER1</i>			Other loci			%homozygosity	
		Hetero.	Hom. alternate	Hom. reference	Hetero.	Hom. alternate	Hom. reference	<i>DICER1</i>	Other loci
Patient 1	Skin Fibroblasts (cultured)	90	9	2	409	231	22	11%	38%
	Normal Brain (frozen)	89	9	3	401	228	28	12%	39%
	Normal Skin (frozen)	79	11	8	364	231	49	19%	43%
	Lymph Node (FFPE)	44	10	14	190	164	83	35%	57%
	PPB Brain Metastasis (10.9y)	86	9	5	384	226	40	14%	41%
	Type II PPB (FFPE)	57	14	14	244	199	60	33%	51%
	PPB Brain Met. (2.1y) (FFPE)	23	14	36	236	201	73	68%	54%
Patient 2	Blood DNA	68	7	4	297	216	40	14%	46%
	Saliva DNA	71	7	2	320	220	28	11%	44%
	Urine DNA	68	7	3	303	210	37	13%	45%
	Left SLCT (FFPE)	55	9	15	262	219	44	30%	50%
	Lung Cysts (FFPE)	10	11	11	36	94	49	69%	80%
Patient 3	Right SLCT (FFPE)	15	29	18	186	208	50	76%	58%
	Thyroid carcinoma (FFPE)	25	20	20	189	201	66	62%	59%
	Blood DNA	41	19	7	507	143	31	39%	26%
	Normal R Kidney #2 (FFPE)	15	15	17	252	146	89	68%	48%
Patient 4	Normal Lung (FFPE)	3	5	7	35	53	86	80%	80%
	Normal R Kidney #1 (FFPE)	19	20	14	291	136	84	64%	43%
	Juvenile Polyps	34	23	10	488	138	42	49%	27%
	Cystic Nephroma (L kidney)	42	19	6	496	142	37	37%	27%
	Mal-devel. lesion (R kidney) (FFPE)	15	22	18	314	138	85	73%	42%
Patient 4	Type I PPB (FFPE)	4	10	11	99	67	115	84%	65%
	Blood DNA	94	9	1	345	209	21	10%	40%
	Reactive Right Lung (FFPE)	66	14	13	233	201	45	29%	51%
	Reactive Left Lung (FFPE)	81	9	6	243	210	54	16%	52%
	Saliva DNA	92	9	3	351	203	15	12%	38%
	Hair DNA	93	9	2	340	203	28	11%	40%
Type II PPB (FFPE)	79	9	9	281	193	42	19%	46%	

Abbreviations: Devel., developmental; Hetero, heterozygous; Hom, homozygous; PPB, pleuropulmonary blastoma; R, right; SLCT, Sertoli-Leydig cell tumor.

Legend: Total number of SNPs and percentage of homozygosity in *DICER1* and other loci. SNPs were called jointly (HaplotypeCaller) using all samples from an individual patient. The percentage of homozygosity represents the percentage of all homozygous SNPs relative to the total number of SNPs of a sample.

TABLE S3

Table S3: Second somatic *DICER1* mutations and the corresponding number of reads obtained in each HaloPlex^{HS} sample

Sample	% Mosaic RNase IIIb mutation	Second somatic <i>DICER1</i> Mutation	*Total Reads	A	C	G	T	Del/Ins	*Total Reads	A	C	G	T	Del/Ins	*Total Reads	A	C	G	T	Del/Ins
Patient 2			Blood DNA					Saliva DNA					Urine DNA							
Left SLCT	44.48%	c.4626_4626delG	2001	0	<i>2001</i>	0	0	<u>0</u>	1130	1	<i>1129</i>	0	0	<u>0</u>	715	0	<i>715</i>	0	0	<u>0</u>
Sinonasal inflammatory polyp	30.55%	c.4458_4458delA	875	0	2	0	873	<u>0</u>	445	0	0	0	445	<u>0</u>	209	0	0	0	290	<u>0</u>
Patient 3			Normal Right Kidney Cortex 1					Normal Right Kidney Cortex 2					Blood DNA							
NCMH	27.73%	c.4651_4652insTGCT	333	0	0	0	333	<u>0</u>	651	1	3	0	647	<u>0</u>	3938	3	3	3	3929	<u>0</u>
Patient 4			Reactive Right Lung					Blood DNA					Saliva DNA							
Type II PPB	37.84%	c.1966C>T	235	<u>0</u>	0	235	0	0	10261	<u>4</u>	0	<i>10256</i>	1	0	3242	<u>0</u>	0	<i>3237</i>	5	0

Abbreviations: Del, deletion; Ins, insertion; Mut, mutant; NCMH, nasal chondromesenchymal hamartoma; PPB, pleuropulmonary blastoma; SLCT, Sertoli-Leydig cell tumor. **Notes:** * duplicate reads removed (HaloPlex^{HS} data); The reference (wild-type) allele at position of interest is *italicised* and the mutant allele is underlined. The Del/Ins indicate the read counts of deletions or insertions identified at the position of interest – counts shown in the fifth columns.

Legend: Second somatic *DICER1* mutations and the corresponding number of reads obtained in each HaloPlex^{HS} sample. The *DICER1* RNase IIIb mosaic mutant allele fraction observed in each respective tumor is indicated in the second column. If contaminating tumor DNA were present, one would expect to see the same ratio of the second somatic mutation in the non-tumor sample(s) as the RNase IIIb mosaic mutation. Since this is not the case, we can conclude that the mosaic mutant allele fraction observed in the non-tumor HaloPlex^{HS} samples is not due to circulating tumor DNA or infiltrating tumor cells.

SUPPLEMENTAL MATERIALS and METHODS

Samples and DNA extraction:

Genomic DNA was extracted from 3 to 5 ml of blood collected in EDTA blood collection tubes from Patients 2, 3 and 4 using the Genra Puregene Blood kit (Qiagen, California, USA) following the manufacturer's instructions. The Genra Puregene Blood kit was also used to extract genomic DNA from saliva and urine samples collected from Patients 1, 2 and 4 using the Oragene DNA OG-250 DNA collection kit (DNA Genotek, Ottawa, Canada) and the BD vacutainer UA preservative collection tubes (BD Diagnostics, NJ, USA) respectively. Genomic DNA was extracted from nail (Patient 1) and hair (Patients 2 and 4) samples using the QIAamp DNA Investigator Kit (Qiagen, Toronto, ON, Canada). DNA was extracted from formalin-fixed paraffin-embedded (FFPE) tissue samples using 3–7 slides of 10 µm thick tumour tissue with the QIAamp DNA FFPE Tissue Kit (Qiagen, Toronto, ON, Canada) according to manufacturer's instructions. DNA from fresh-frozen tumour tissue was extracted using the Qiagen AllPrep DNA/RNA Mini Kit (QIAGEN, Toronto, ON, Canada).

Sanger sequencing and Multiplex Ligation-based Probe Amplification assay (MLPA):

Screening for germ-line *DICER1* mutations was conducted on DNA extracted from blood and/or saliva by PCR and Sanger sequencing of the region of interest.^{3; 4} We screened for large deletions or duplications using our MLPA assay.³ The *DICER1* RNase domains were screened for “hotspot” mutations by conventional Sanger sequencing using previously published primers.^{4; 5} DNA from any FFPE sample in which a mutation was found was extracted twice independently and the PCR was repeated at least twice from each independent extraction using QIAGEN HotStarTaq, 10 mM dNTP and 10× PCR buffer reagents with 1.4 µl of 20 µM primers in a 25-µl PCR reaction.

PCR amplification of the regions of interest was done using a Touchdown PCR program (adapted from Witkowski et al., 2013⁵) with the following thermocycler parameters: 95°C for 15 min, (95°C for 20 s, 63°C for 20 s, 72°C for 30 s) for 8 cycles, (94°C for 20 s, 58°C for 20 s, 72°C for 30 s) for 28 cycles, 72°C for 7 min.

Primer pairs for this purpose were designed using Primer3 (<http://frodo.wi.mit.edu/>) and were run through a UCSC in silico PCR software (<http://genome.ucsc.edu/cgi-bin/hgPcr?command=start>) to ensure yield of a single, specific product. Visualization of the PCR products was done in a 1.5% agarose gel with GelRed (Biotiom), and sequencing was performed by the McGill University and Genome Quebec Innovation Centre (MUGQIC).

Fluidigm Access Array:

The Fluidigm Access Array System involves an array-based PCR amplification of a specific region of interest (target enrichment). For our purpose, the exons of *DICER1*, all exon-intron boundaries and the 3'UTR were selectively targeted as previously described.⁶ Next-generation sequencing was carried out using the Illumina MiSeq (McGill University and Genome Quebec Innovation Centre (MUGQIC)) and the Integrative Genomics Viewer software (IGV version 2.3; <http://www.broadinstitute.org/igv/>) was used to analyse the dataset. All additional variants identified were confirmed with Sanger sequencing.

Supplemental References:

1. Nur, S., Badr, R., Sandoval, C., Brudniki, A., and Yeh, A. (2007). Syndromic presentation of a pleuropulmonary blastoma associated with congenital cystic adenomatoid malformation. A case report. *J Pediatr Surg* 42, 1772-1775.
2. Ramasubramanian, A., Correa, Z.M., Augsburger, J.J., Sisk, R.A., and Plager, D.A. (2013). Medulloepithelioma in DICER1 syndrome treated with resection. *Eye* 27, 896-897.
3. Sabbaghian, N., Srivastava, A., Hamel, N., Plourde, F., Gajtko-Metera, M., Niedziela, M., and Foulkes, W.D. (2014). Germ-line deletion in DICER1 revealed by a novel MLPA assay using synthetic oligonucleotides. *Eur J Hum Genet* 22, 564-567.
4. Wu, M.K., Sabbaghian, N., Xu, B., Addidou-Kalucki, S., Bernard, C., Zou, D., Reeve, A.E., Eccles, M.R., Cole, C., Choong, C.S., et al. (2013). Biallelic DICER1 mutations occur in Wilms tumours. *J Pathol* 230, 154-164.
5. Witkowski, L., Mattina, J., Schonberger, S., Murray, M.J., Choong, C.S., Huntsman, D.G., Reis-Filho, J.S., McCluggage, W.G., Nicholson, J.C., Coleman, N., et al. (2013). DICER1 hotspot mutations in non-epithelial gonadal tumours. *Br J Cancer* 109, 2744-2750.
6. de Kock, L., Sabbaghian, N., Plourde, F., Srivastava, A., Weber, E., Bouron-Dal Soglio, D., Hamel, N., Choi, J.H., Park, S.H., Deal, C.L., et al. (2014). Pituitary blastoma: a pathognomonic feature of germ-line DICER1 mutations. *Acta Neuropathol* 128, 111-122.