

Supplementary Information – Methods

Ethical approval was obtained from North Somerset & South Bristol Research Ethics Committee (reference 09/H0106/72 for data collection and reference 09/H0106/80 for sample collection and data analysis).

Array Design

Arrays were designed by Roche NimbleGen. Overlapping target regions were merged and short target regions extended to 250bp to increase capture efficiency. Unique probes were around 75bp in length. Repetitive regions were excluded in the design to avoid off-target capture.

DNA library preparation

DNA was extracted using a QIAamp DNA Blood Mini Kit (Qiagen). Each sample was quantified and the DNA quality checked using a Nanodrop (ThermoScientific) prior to use.

Sample libraries were constructed following Illumina multiplexing sample preparation protocol (Illumina Catalog # PE-930-1002 Feb. 2010) with personal modifications.

5µg of genomic DNA was quantified on a Qubit Fluorometer (Invitrogen) and then sonicated using a Biorupter UCD-200 (Diagenode) to create DNA fragments between 100 and 600bp in length. Fragments were blunt ended and 5' ends phosphorylated. Custom adapters (Transcriptomics Unit, University of Bristol) were ligated (NEBNext DNA sample prep reagent set 1-E6000S). Each of the adapters contained a bar code, unique for each sample on the array.

Ligated products were purified on a 2% agarose gel (120 V, 75min). 350bp fragments (SYBR Safe DNA Gel Stain, Invitrogen; Dark Reader Transilluminator) were excised and purified on column (QIAquick Gel Extraction Kit, Qiagen) prior to pre-capture PCR. 10 amplification cycles were performed, PCR products were column purified (QIAquick PCR Purification Kit, Qiagen), eluted in 50µl of PCR-grade H₂O and quality validated using an Agilent 2100 Bioanalyser (Agilent Technologies). Two NimbleGen HX12 arrays were used. 300ng of each DNA library was hybridised on a NimbleGen HX12 platform. Hybridization end elution was prepared using a NimbleGen 454 optimized sequence capture arrays protocol (confidential with personal modifications).

Hybridisation, Capture and Sequencing

Hybridization enhancing oligos (HE) were designed to correspond to the ligated adapters, containing barcodes, and primers. After 70 hours of hybridization, the array was washed and the 12 samples were eluted together and post-capture PCR was performed. The final product (containing all 12 samples) was quantified using a Qubit Fluorometer (Invitrogen); quality checked using a Nanodrop (ThermoScientific) and validated on an Agilent 2100 Bioanalyser (DNA 1000 chip, Agilent). 110bp paired-end sequencing (Illumina GAIIx), base calling and demultiplexing were performed in the Transcriptomics Unit (University of Bristol).

Bioinformatics Pipeline

Data analysis was performed using CLC Genomics Workbench v 4.9 – 5.0 (CLC Bio). dbSNP (build 132) was annotated onto the reference sequence (hg19/GRch37) prior to mapping, as were capture targets corresponding to probes designed by Roche NimbleGen. Duplicate

reads were removed and reads were quality trimmed (default parameters). Unique reads were then mapped against the human genome (hg19/GRCh37) with standard mapping parameters (see supp table 1), paired distance was 80-500bp and reads matched to multiple locations were ignored (i.e. when a read matches equally well on two or more regions of the reference sequence). Detection of single nucleotide polymorphisms (SNPs) and short insertion/deletion polymorphisms (indels) detection was performed using listed parameters (supp table 2). Splice site effect prediction was performed using Genomics Gateway (version 2.0 beta5) identifying any SNPs that fell within 2 base pairs of an intron-exon boundary

Supplementary Information – Results

Supplementary Table 1

mismatch cost	2
insertion cost	3
deletion cost	3
length fraction	0.5
similarity	0.8
non-specific matches	ignore

Supplementary Table 2

	SNP detection	DIP detection
window length	10	-
maximum number of gaps and mismatches	2	-
minimum average quality of surrounding base	15	-
minimum quality of central base	20	-
minimum coverage	4	4
minimum variant frequency(when coverage 4-7)	50%	50%
minimum variant frequency(when coverage 8-19)	25%	25%
minimum variant frequency(when coverage >20)	15%	15%
maximum expected variations (ploidy)	2	2

Supplementary Table 3.

Gene	Variant	Patient	dbSNP ref# (build 135)	Minor Allele Frequency (%)	PolyPhen	MutPred - general score
COL4A3	Asp1269Glu	36	rs57611801	4.1	unknown	0.135
COL4A4	Gly545Ala	11	rs1800516	2.7	probably damaging	0.643
COL4A4	Met1552Ile	16	rs77104306	1.4	unknown	0.44
COL4A4	Ala1078Val	16	rs79143859	1.4	unknown	0.51
COL4A4	Ile967Val	16	rs80243096	1.6	unknown	0.269
INF2	Pro1096Ser	25	rs34251364	4.5	unknown	0.228
LAMB2	Gly914Arg	2, 29	rs35713889	3.0	probably damaging	0.763
LAMB2	Ala1765Thr	36	rs74951356	1.9	benign	0.335
PLCE1	Ser469Thr	26	rs17508082	2.3	benign	0.45
SMARCAL1	Ser315Arg	1	rs2066522	3.0	benign	0.084