**Supplementary methods**

***What if we would use a diagnostic multi-cancer gene panel for opportunistic screening? A study in 2,090 Dutch familial cancer patients***

L.F. Johansson1, K.K. van Dijk-Bos1, A.H. van der Hout1, A.P. Knopperts1, B. Leegte1, P.C. van den Akker1, K. Kok1, I.M. van Langen1 , M.A. Swertz1, R. K. Weersma2, R.J. Sinke1, B. Sikkema-Raddatz1, H. Westers1\*, R.H. Sijmons1\*

1 University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands

2 University of Groningen, University Medical Center Groningen, Department of Gastroenterology, Groningen, the Netherlands

\* these authors contributed equally to the paper

## **Selection of control samples: Lifelines Deep cohort**

The LifeLines Deep (LLD) cohort consists of individuals from the general population in the northern Netherlands [Tigchelaar, 2015]. For 1000 LLD participants, exome sequencing data was available. A subset of these participants (508 individuals) had filled in a questionnaire on irritable bowel syndrome (IBS), and 172 of them were considered to have self-reported IBS according to ROME III criteria [Longstreth, 2006]. We excluded these 172 from our control set because of the similarity in symptoms between IBS and colorectal cancer [Hsiao, 2014]. This left 828 LLD samples that could be used as controls. However, possible cancer phenotypes of Genome of the Netherlands (GoNL) or LLD participants were unknown to the researchers and cannot therefore be excluded.

## **Alignment procedure and variant calling**

Using the Burrows-Wheeler Aligner (BWA) v0.7.12 [Li and Durbin, 2010], sequence data were first aligned to the human reference genome build b37 as released by the 1000 Genomes Project [1000 Genomes Consortium, 2010]. Duplicates were marked using Sambamba v0.6.3 [Tarasov et al., 2015]. These steps were part of a production pipeline created in MOLGENIS compute [Byelas et al., 2013]. A Browser Extendible Data (BED) file was used to specify regions of interest by listing the chromosome, start position, end position and gene for all targets in a tab-delimited file. The Binary Alignment Map (BAM) files created [Li et al., 2009], together with the relevant BED file, were used as input for two CNV calling algorithms: CoNVaDING [Johansson et al., 2016] and XHMM [Fromer et al., 2012]. SNVs and indels were called using the GATK Haplotype Caller v3.5. Variants in the GoNL and LLD datasets for targeted regions were retrieved as described above.

The GoNL project consists of 769 individuals from 250 families of Dutch ancestry analyzed using WGS, with a 12x coverage on average [The genome of the Netherlands consortium, 2014; Kloosterman et al., 2015; Francioli et al., 2015]. High confidence calls were made publicly available (<http://www.nlgenome.nl/>). Variant calls for SNPs and indels (release 5) and structural variants (SVs) (release 6) were included in our analysis when they overlapped with one of the genes included in our gene panels. For SNP and indel analyses only, calls from 498 unrelated parents are used, while the SV list is an aggregate of calls made in all participants.

For the LLD samples, read alignment and variant calling was performed as described in Rivas et al., 2018 [Rivas et al., 2018]. The average coverage of the samples was 89x (sd 4x), and 89% of the exons, on average, were covered by at least 20 reads.

## **CNV calling**

For both gene panels, a folder was created in which the BAM files of all samples were placed for the respective capture design. For both folders, CoNVaDING analysis was performed. If a parameter is not named in the text below, default settings were used. To start we used the StartWithBam mode, using the rmdup and useSampleAsControl parameters, to produce a sample group and a control group consisting of the same samples. A BED file was used to define the targets and genes per panel, as described above. Subsequent analysis was divided into two steps: a preliminary investigation was performed to clean up the control group and a second analysis was performed for prediction using the curated control group. To clean up the control group, we used the StartWithMatchScore mode using the sample group as inputDir and the control group as controlsDir. In the third analysis step, we used StartWithBestScore to create a shortlist for each sample containing all CNV calls passing QC based on the control group selected for this sample. Samples with a call on the shortlist were removed from their specific control group to create curated control groups. All samples were reanalyzed using the curated control groups, starting from the StartWithMatchScore step. The two final steps of the workflow, GenerateTargetQcList and CreateFinalList, were also performed to include quality information from all analyzed samples in the analysis. Sample quality information was obtained during the StartWithBam step. A sample passed QC at a sample coefficient of variation < 0.09.

In addition, all samples were analyzed using XHMM. For each design, all samples were analyzed in batch using the same settings we used in our previous project [Johansson et al., 2016].

A CNV call was made by one of three routes. The first yielded all calls present on the CoNVaDING final list and made by XHMM in samples that passed CoNVaDING sample QC. The second route contained calls that were made by XHMM only, but were present in the CoNVaDING longlist in samples that passed CoNVaDING sample QC. The XHMM calls were retained if a CoNVaDING call was present on the longlist but filtered due to low quality, if CoNVaDING ratios and Z-scores were very close to the threshold, or for large calls if CoNVaDING ratios were consistently lowered. In the third route, calls made only by CoNVaDING in samples passing CoNVaDING sample QC were retained if the selected control group was normally distributed for at least one target in the call and if the ratios and Z-scores were well outside the thresholds. Apart from these official calls, failed samples were looked at manually. Here we considered calls that we judged to have a high chance of being true because they entailed a deletion or duplication of a high number of targets of a single gene or they gave a call in published deletion regions.

## ***POLE* and *POLD1* sequencing**

Genomic DNA was PCR amplified. PCR for *POLE* was performed with (POLE\_Fw) 5’-GGTGCCTGTTAGGAACTTGC-3’and (POLE\_Rv) 5‘-TAGCTCCACGGGATCATAGC-3’ primers. PCR for *POLD1* was performed with (POLD1\_Fw) 5’-GGAGTACAAGCTCCGCTCCT-3’ and (POLD1\_Rv) 5’-GCGTCAGGTAAGGTCACAGG-3’ primers. A standard amplification program was used: 5 minutes 94 °C, 30 cycles 45 seconds 94 °C, 45 seconds 60°C, 60 seconds 72 °C, 10 minutes 72 °C. Next, 10 ng PCR product was used for Sanger sequencing using an ABI 3730. Analysis of variants was performed with Mutation surveyor software (Mutation surveyor, version 3.23, Softgenetics LLC, Pennsylvania).

## **References**

[1] Tigchelaar EF, Zhernakova A, Dekens JAM, et al. Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. *BMJ Open*. 2015;5(8):e006772.

[2] Longstreth GF, Thompson WG, Chey WD, et al. “Functional Bowel Disorders,” *Gastroenterology*. 2006;130(5):1480-1491.

[3] Hsiao CW, Huang WY, Ke TW, et al. Association between irritable bowel syndrome and colorectal cancer: A nationwide population-based study. *Eur. J. Intern. Med.* 2015;25(1):82-86.

[4] Li H, Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;26**:**589–595.

[5] The 1000 Genomes Project Consortium. A map of human genome variation from population scale sequencing. *Nature*. 2010;467(7319):1061-1073.

[6] Tarasov A, Vilella A J, Cuppen E et al. Sambamba: Fast processing of NGS alignment formats. *Bioinformatics* 2015;31**:**2032-2034.

[7] Byelas H, Dijkstra M, Neerincx P, et al. Scaling bio-analyses from computational clusters to grids. *Proceedings of the 5th International Workshop on Science Gateways, Zurich, Switzerland.* 2013.

[8] Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078-2079

[9] Johansson LF, van Dijk F, de Boer EN, et al*.* CoNVaDING: Single Exon Variation Detection in Targeted NGS Data. *Hum Mutat*. 2016;37**:**457–464.

[10] Fromer M, Moran JL, Chambert K, et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am . Hum Genet*. 2012;91:597–607.

[11] The Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet*. 2014;46**:**818-825.

[12] Kloosterman WP, Francioli LC, Hormozdiari F, *et al.* Characteristics of de novo structural changes in the human genome. *Genome Res.* **25,** 792–801 (2015).

[13] Francioli LC, Polak, PP, Koren A, *et al.* Genome-wide patterns and properties of de novo mutations in humans. *Nat. Genet.* **47,** 822–826 (2015).

[14] Rivas MA, Avia BE, Koskela J, et al. Insights into the genetic epidemiology of Crohn’s and rare diseases in the Ashkenazi Jewish population. *Plos Genetics*. 2018;14(5):e1007329.