UVIC UNIVERSITAT DE VIC

Master of Science in Omics Data Analysis

Master Thesis

Copy number variations of colorectal cancer by whole exome sequencing data

by

Ana Maria Corraliza Márquez

Supervisor: Victor Moreno, Unitat de Biomarcadors i Susceptibilitat, Institut Català d'Oncologia

Co-supervisor: M. Luz Calle Rosingana, Systems Biology Department, University of Vic

Department of Systems Biology

University of Vic – Central University of Catalonia

September 24, 2014

Abstract

Colorectal cancer (CRC) is the third most common cancer and the fourth leading cause of cancer death worldwide. About 85% of the cases of CRC are known to have chromosomal instability, an allelic imbalance at several chromosomal loci, and chromosome amplification and translocation. The aim of this study is to determine the recurrent copy number variant (CNV) regions present in stage II of CRC through whole exome sequencing, a rapidly developing targeted next-generation sequencing (NGS) technology that provides an accurate alternative approach for accessing genomic variations. 42 normal-tumor paired samples were sequenced by Illumina Genome Analyzer. Data was analyzed with Varscan2 and segmentation was performed with R package R-GADA. Summary of the segments across all samples was performed and the result was overlapped with DEG data of the same samples from a previous study in the group [1]. Major and more recurrent segments of CNV were: gain of chromosome 7pq(13%), 13q(31%) and 20q(75%) and loss of 8p(25%), 17p(23%), and 18pq(27%). This results are coincident with the known literature of CNV in CRC or other cancers, but our methodology should be validated by array comparative genomic hybridisation (aCGH) profiling, which is currently the gold standard for genetic diagnosis of CNV.

Table of Contents

List of Tables

List of Figures

1

Introduction

Colorectal cancer (CRC) is the third most common cancer and the fouth leading cause of cancer death worldwide, with over 1.2 million new cancer cases and 608.700 deaths estimated in 2008[2].

CRC incidence rates are increasing in areas that historically were at low risk, including Spain, and a number of countries within Eastern Europe and Asia[3]. According to *Cancer Statistics, 2014*, CRC is estimated to account for 5.4% and 8.5% of total cancer incidence and death, respectively, in the United States in 2014[4].

Tumor formation in the colon-rectum begins as a benign adenomatous polyp, which develops into an advanced adenoma with high-grade dysplasia and then progresses to an invasive cancer[5]. This process takes years to decades and it is influenced by an interaction between intrinsic and extrinsic factors, including gender, age, diet (e.g intake of red meat and excessive alcohol), comorbidities (e.g. obesity, inflammatory bowel diseases)[6], diabetes[7] and lifestyle (e.g physical activity, smoking)[8].

Recent findings also show the implication of gut microbia in CRC development: *Fusobacterium* spp was found to be enriched in a subset of human colon adenomas[9] and has been shown to potentiate tumorigenesis in mouse models and cell culture experiments by stimulating inflammation[10].

Such interactions at different levels promote the acquisition of genetic and epigenetic abnormalities, that contribute to the aberrant activation of proto-oncogenes (e.g KRAS, PIK3XA, BRAF) and inactivation of tumor-supressor genes (e.g APC, TP53)[11].

The majority of CRCs are sporadic (70%-80%) and only a small proportion of cases are due to inherited forms, either familial adenomatous polyposis (less than 1%), non-polyposis hereditary CRC or Lynch syndrome (2%-5%) or MYH-gene associated polyposis (< 1%)[12].

Up to one third of CRC cases (20-25%) are estimated to be caused by hereditary components, but disease-predisposing mutations account for less than a 5% of the cases[13].

Recently, copy number variation (CNV) has been recognized as one of the most important genomic alteration that play a role in cancer pathogenesis[14]. In addition, most sporadic cases of CRC (about 85%) are known to have chromosomal instability, an allelic imbalance at several chromosomal loci, and chromosome amplification and translocation[15]. Chromosomal aberrations

frequently reported in CRC are 7pq, 8p 13q and 20q gains, and 4pq, 5q, 8p , 15q, 17p and 18q losses[16–18].

Array comparative genome hybridation (aCGH) has been widely used to identify copy number variations in genomes[19–22]. However, aCGH is expensive and has limited resolution and accuracy[23]. Nowadays, rapidly developing next-generation sequencing (NGS) technologies provide an accurate alternative approach for accessing genomic variations. The speed, quality and affordability give NGS a significant advantage over microarrays[24]. In practice, many recent disease studies have chosen high-depth targeted exome sequencing, also known as whole exome sequence (WES) technology[25,26]. The key reasons for choosing this methodology are the lower cost and higher coverage as compared to whole-genome sequencing, its effectiveness and the interpretability of a variant's effect on gene product[27,28]. Notwithstanding the advantages of NGS, its use for CNV identification has been limited by a lack of available and effective statistical and methodological approaches. Furthermore, detecting CNV from exome sequencing is challenging because of the noncontiguous nature of the captured exons.

In this study, our main goal was to determine the most common CNV regions present in 42 samples of Colorectal cancer that were sequenced by WES technology.

2

Material and Methods

2.1 Samples obtainment

A sub-set of 42 paired adjacent normal and tumor tissues (84 samples) belonging to a previously described set of 98 CRC patients with stage II and microsatellite stable tumors (GSE44076)[29] were included in this work (Supplementary Table A.1). All patients were recruited at the Bellvitge University Hospital (Spain) and the Ethics Committee approved the protocol. Written informed consent from patients and healthy donors was required for inclusion in this study. DNA was extracted using the phenol-chloroform protocol. To ensure that normal and tumor tissues were paired, dynamic arrays were used to genotype the 84 samples using 13 SNPs (Supplementary Table A.2). Indeed, all 42 normal correctly matched with their corresponding tumor (Supplementary figure B.1).

2.2 Exome sequencing and alignment

Exomic DNA (3 μgr) from the set of 42 normal-tumor paired samples was sequenced in CNAG (National Center of Genomic Analysis) by Illumina Genome Analyzer. Exome capture was performed using Sure Select kit (Agilent). Tumoral exomes were sequenced at 60X coverage and exomes from normal tissue were sequenced at 40X. FastQ software was used to assess sequences quality (http://ww w.bioinformatics.bbsrc.ac.uk/projects/fastqc). All samples displayed good quality, and accordingly all of them were considered for further analysis. Bowtie 2.0 software was used to align sequences over the human reference genome HG19[30]. To refine data, reads unmapped, reads with mate unmapped, not primary alignments, and reads that were PCR or optical duplicates were rejected (http://picard.sourceforge.net/).

2.3 Data analysis

There are five approaches to look for CNV in NGS data: paired end mapping, split read, read depth, assembly-based and combinatorial approach[24,31]. In this study a read depth

(RD) based approach was used, as it is known to be particularly effective for exome data considering that it does not rely on sequencing into or near the CNV breakpoints[31].

The underlying hypothesis of RD methods is that the depth of coverage (DOC) in a genomic region is correlated with the copy number of the region; a loss of copy number should have a lower intensity than expected[32].

In this case, there were case/control matched pair of samples per each patient. Therefore, the CNV detection was leveraged on the matched control sample, which served as a reference genome.

Supplementary Table A.3 shows alternative software to detect CNV in WES data. In this study only autosomal chromosomes were analyzed.

Normalization and estimation of copynumber:

This step was performed with Samtools and Varscan2[33].

Samtools *mpileup* command was run on normal and tumor BAM files in order to get mpileup files needed for Varscan2 protocol. Secondly, Varscan2 *copynumber* command was run on normal and tumor mpileup output. Here, the minimum and maximum segment size allowed was 100 and 500 pair bases respectively (by default 10 and 100 respectively). With this step the raw copynumber regions are obtained, with chromosome, start position, stop position, and relative copy number.

The relative copy number in tumor is computed from the log-base-2 of the ratio of tumor depth to normal depth. VarScan reports contiguous regions with similar read depths. To reduce noise from spurious differences at low coverage, the minimum coverage requirement was set to 20.

Region boundaries are determined by:

1. gaps of 2 consecutive positions that have not the minimum coverage, or
2. change-points at which the ratio of normal depth to tumor depth changes significantly ($p<0.05$ with Fisher's Exact Test of read depths at this position compared to those at previous position)

Finally, *copyCaller* command was performed to adjust raw copynumber (log_2) values for GC content. This step was made considering the known bias produced by the GC-content in high-throughput sequencing[34]. Also, the raw copynumber data was recentered because neutral segments were not on the ($log_2 = 0$) axis due to the different depth of coverage in normal and tumor samples (40X and 60X, respectively). The code used in this step can be seen in Appendix C.1

Segmentation:

GC-corrected log_2 values underwent segmentation analysis with the R-GADA package[35] to produce segmented calls delineated by significant change-points. The breakpoint identification of this R package is the sparse Bayesian learning[36] followed by a Backward elimination procedure[35]. The variables used in this steps are:

aAlpha = 0.8 Controls the maximum breakpoint sparseness (sensitivity)

T = 5 Adjusts the level of FDR

MinSegLen=3 Minimum number of altered probes in each segment

With the values used, our goal is to achieve a low sensibility and a low FDR.
The code used in this step is shown in Appendix C.2.

2.4 Recurrent CNV regions

The code used to find the recurrent CNV regions is shown in Appendix C.3. In order to make the values more easily understandable and visual, a summary plot was made with the segments present in the 42 samples along the chromosomes analyzed. Segments were considered as a real gain or loss when the log_2 adjusted ratio was higher than 0.5 or lower than -0.5, respectively.

The functions made to do this summary are based in the R-GADA package[35].

2.5 Collection of Genes in CNV regions

The determination of the Genes in each region was made by the BiomaRt R package. The genes belonging to the regions that were present in at least a 15% of the samples were queried. Due to the amount of genes present in each region, and in order to have a better idea of the cellular functions affected, a search of the GO terms of each gene was also performed by this package.

3

Results

After performing our pipeline, there were 10581 segments in total, with a mean of 251.92 segments per sample. The length of the segments had a mean of 1060780 pair bases (pb), from a minimum of 269pb to a maximum of 190910284 pb (Table3.1). A majority of segments were considered as lost segments or delections (6556, 61.96 %). Not surprisingly, this percentage was not maintained in all chromosomes: in chromosome 7, 13 and 20 there is a preponderance of won segments or duplications (62.3%, 67.7% and 79.4 %, respectively; Table 3.2).

Segments	
Total segments	10581
Length(pb)	
Minimum	269
Mean	1060780
Maximum	153345286
SD	3831224
Quantiles	
25%	20701
50%	125094
75%	614956

Table 3.1: **Descriptive data of the Segments produced by our pipeline**

The segments across all samples were summarized in Figure 3.1. Major alterations across all samples are: gain of chromosome 7pq(13%), 13q(31%) and 20q(75%) and loss of 8p(25%), 17p(23%), and 18pq(27%) (Figure 3.1). Most stable chromosomes were 2 and 22 (Figure 3.1, Figure B.2). It is also clear that the majority of segments are considered as lost and that the gain segments are basically on chromosome 7, 8, 13, 16 and 20 (Figure 3.2). On the other hand, lost segments are distributed more equally across the genome, with the remarkable lost of chromosome 8p, 17p, 18pq, and 21q (Figure 3.3).

Genes in the regions of lost and won segments shared across 15% of the samples are shown in Appendix D.1 and D.2. These genes were compared with the gene expression analysis performed

Figure 3.1: **Summary of recurrent CNV**. The figure shows the segments of loss and gain copy number (in blue and red, respectively) across all samples. The height of the segment represents the frequency across all samples, being the represented lines a 25%, 50%, 75% and 100% of them. It can be observed that the major changes are seen in chromosome 7, 8, 13, 17, 18 and 20.

Figure 3.2: **Summary of won CNV**. The figure shows only the segments considered as won across all the samples. The height of the segment represents the frequency across all samples, being the represented lines a 25%, 50%, 75% and 100% of them. It can be seen that the major changes are in chromosome 7, 8, 13 and 20.

Figure 3.3: **Summary of lost CNV**. In the figure are only represented the segments considered as lost across all the samples. The height of the segment represents the frequency across all samples, being the represented lines a 25%, 50%, 75% and 100% of them. It can be seen that the major changes are located in chromosome 8, 17 and 18. Nevertheless, changes are distributed more unifomly across all the chromosomes, unlike segments considered as won, which are concentrated in chromosome 7, 8, 13 and 20 (Figure 3.2).

Chromosome	**Segments**	**Gains**(%)	**Losses**(%)
chr 1	1100	362 (32.9%)	738 (67.1%)
chr 2	911	362 (39.7%)	549 (60.3%)
chr 3	678	220 (32.4%)	458 (67.6%)
chr 4	678	139 (20.5%)	539 (79.5%)
chr 5	675	205 (30.4%)	470 (69.6%)
chr 6	613	205 (33.4%)	408 (66.6%)
chr 7	581	362 (62.3%)	219 (37.7%)
chr 8	602	241 (40%)	361 (60%)
chr 9	328	102 (31.1%)	226 (68.9%)
chr 10	474	142 (30%)	332 (70%)
chr 11	560	196 (35%)	364 (65%)
chr 12	559	244 (43.6%)	315 (56.4%)
chr 13	394	267 (67.7%)	127 (32.3%)
chr 14	388	92 (23.7%)	296 (76.3%)
chr 15	320	80 (25%)	240 (75%)
chr 16	237	116 (48.9%)	121 (51.1%)
chr 17	347	129 (37.2%)	218 (62.8%)
chr 18	297	26 (8.75%)	271 (91.75%)
chr 19	236	138 (58.5%)	98 (41.5%)
chr 20	412	327 (79.4%)	85 (20.6%)
chr 21	83	26 (31.3%)	57 (68.7%)
chr 22	108	44 (40.7%)	64 (59.3%)
Total	**105181**	**4025** (38.05%)	**6556** (61.96%)

Table 3.2: **Number of segments in all samples per chromosome.**

for the same samples previously in the group[1]. Genes matching lost CNV and loss of function and vice versa shared across 15% of the samples are shown in Table 3.3. GO terms analysis was performed for the genes won and lost copy number regions. It has been observed that the pathways affected by the lost regions are associated with the positive regulation of apoptotic processes, like tumor necrosis factor (TNF), p53 and Toll-like receptors. Otherwise, the gain copy number regions have affected different pathways, related with cytoskeleton organization, positive regulation of mitosis and inflammatory response, and telomere maintenance (data not shown).

Gene	Molecular Function	(% of samples)
	Genes with CNV loss	
GFMB	inhibition of proliferation of tumor cells	11.9%
DEFA5	innate immune response	26.2%
	Genes with CNV won	
FRY	microtubule organizing center	31%
WASF3	actin binding	32%
KATNAL1	microtubule binding	31%
LHFP	DNA binding	28.5%
MAFB	DNA, protein and transcription factor binding	71.4%
PI3	endopeptidase, serine-type endopeptidase and peptidase inhibitor activity	70.2%
TSHZ2	sequence-specific DNA binding transcription factor activity, protein binding, sequence-specific DNA binding and metal ion and DNA binding	69.4%
SOX18	RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in positive regulation of transcription, sequence-specific DNA binding transcription factor activity	68.9%
THBD	transmembrane signaling receptor activity, calcium ion binding, protein binding, carbohydrate binding	35.3%
CD93	complement component C1q binding, receptor activity, calcium ion binding, protein binding, carbohydrate binding, receptor activity	35.3%
PRNP	copper ion binding, protein binding, microtubule binding, tubulin binding, identical protein binding, microtubule binding	21.6%
INMT	thioether S-methyltransferase activity, methyltransferase activity, amine N-methyltransferase activity	16.7%
AEBP1	DNA binding, sequence-specific DNA binding transcription factor activity, transcription corepressor activity, carboxypeptidase activity, metallocarboxypeptidase activity	16.7%
IGFBP3	fibronectin binding, protein binding, insulin-like growth factor binding, protein tyrosine phosphatase activator activity, insulin-like growth factor I binding	16.6%
ELN	extracellular matrix structural constituent, protein binding, extracellular matrix constituent conferring elasticity, extracellular matrix binding	15.1%
PHF20L1	protein binding, zinc ion binding	22%
TRIB1	protein kinase inhibitor activity, transcription factor binding, transferase activity, transferring phosphorus-containing groups	18.4%
HAS2	hyaluronan synthesis, regulation of cell adhesion, migration and differentiation	17.5%

Table 3.3: **Table with the genes overlapping with the DEG found previously in our group**[1]. The segments of our study considered for the overlapping with this genes are the ones shared across at least one 15% of the samples.

4

Discussion

It has long been considered that genomic instability is a driver of CRC, by facilitating the acquisition of multiple tumor-associated aberrant gene expression. The most common type of genomic instability in CRC is chromosomal instability, which causes numerous changes in chromosomal structure and copy number[13]. In this study we have used exome sequencing technology, a rapidly developing technology that gives advantage over microarrays[20], to describe the major CRC copy number changes at chromosome level. We have found copy number gain of chromosome 7pq, 13q and 20q and loss of 8q, 17p, and 18pq.

Gain of chromosome arm 20q was the most frequent site of increased copy number in our samples (75%). 20q gain was a previously well-defined arm-level change in colorectal carcinoma[14,25–28]. Frequent gains of chromosome arm 20q have been observed in several other human carcinomas, including bladder cacer[29], ovarian cancer[37] and pancreatic cancer[38]. Moreover, gains in chromosome 13q and 20q have been associated with adenoma-to-carcinoma progression[14] and it is associated with a poor prognosis and metastasis in CRC and other cancer types[27,39,40].

The fact that amplification of substantial parts of 20q is associated with progression of colorectal adenoma to carcinoma implies that increased DNA copy numbers of one or more genes must have functional consequences, most likely caused by aberrant mRNA and protein expression. Many genes have been described as putative oncogenes at this amplicon, i.e., C20orf24, AURKA, HNF4A, TPX2 and ZNF217[28,41,42]. All these genes are found in the region 20q amplified in the majority of our samples (Figure 3.2). AURKA is specially related with transformation and tumorigenesis, and it is known to directly interact with TPX2, which stimulates AURKA activity[41,43]. As we have seen, amplification of chromosome 20 involves the whole q-arm of 36Mb. Given the known functional interaction of these two genes, coamplification appears to be a possible mechanism for achieving a selection advantage that could be present in other genes in the amplicon and lead to adenoma to carcinoma progression.

Gain of chromosome 7 is also common in CRC[44]. Furthermore, it is correlated with prostate cancer metastasis[45], and is known to be an idiosyncratic abnormality of papillary renal cell carcinoma[46]. Chromosome arm 7p12 contains the EGFR proto-oncogene, and, the MET proto-oncogene maps to 7q31.1. Both are some of the genes that are thought to be part of the malignant pattern of chromosome 7 aberration.

Chromosome 18q deletion has been described as a negative prognostic factor in CRC[47] and has been identified in as many as 70% of primary colorectal tumors, particularly in advanced stages[5]. In our case, it was found on a 27% of the samples. DPC4, SMAD2, SMAD4, Cables and DCC are tumor suppressor genes reported in 18q and known to be lost, mutated and/or silenced in CRC[48–50]. They are all related with cell growth, differentiation and apoptosis. Together with 20q amplification, 18q loss is known to discriminate between hereditary non-ploypposis colorectal cancers (Lynch syndrome and familial colorectal cancer)[51].

The short arm of chromosome 17 is long known to be frequently lost in CRC[52]. In our results, we found 17p lost in a 23% of our samples. Loss of chromosomal arm 17p could lead to a depletion of the TP53 gene, a tumor suppressor gene that plays a major role in apoptosis, genomic stability, and inhibition of angiogenesis[53].

8p deletion (25%) and 8q gains (13%) are also major chromosomal alterations found in our samples. They are usually found in prostate cancer, and are associated with tumor progression and poor prognosis[54]. These alterations are not yet described in colorectal carcinoma.

Overlaping differentially expressed genes of the previous study realized with the same samples in our group[1] showed a 6% of overlapping DEG with the genes found in the regions with CNV present, at least, in 15% of our samples (Table 3.3). These genes have molecular functions related with cancer. The low percentage of overlapping genes could be related with the different causes of deregulation in genes, including mutations and small indels.

In all, we have seen that our findings are consistent with the literature, and that our most recurrent CNV are previously known for CRC or other cancers like breast or prostate. Hence, our methodology seems to be solid. Nevertheless, these findings should be validated in the same samples with other methodologies, to define its accuracy and validity.

Bibliography

[1] Rebeca Sanz-Pamplona, Antoni Berenguer, David Cordero, David Mollevi, Marta Crous-Bou, Xavier Sole, Laia Pare-Brunet, Elisabet Guino, Ramon Salazar, Cristina Santos, Javier de Oca, Xavier Sanjuan, Francisco Rodriguez-Moranta, and Victor Moreno.
Aberrant gene expression in mucosa adjacent to tumor reveals a molecular crosstalk in colon cancer.
Molecular Cancer, 13(1):46, 2014.

[2] Ahmedin Jemal, Freddie Bray, Melissa M. Center, Jacques Ferlay, Elizabeth Ward, and David Forman.
Global cancer statistics.
CA: A Cancer Journal for Clinicians, 61(2):69–90, 2011.

[3] A Center MM, Jemal, RA Smith, and E. Ward.
Worldwide variations in colorectal cancer.
CA Cancer J Clin., 59(6):366–78, Nov-Dec 2009.

[4] Rebecca Siegel, Jiemin Ma, Zhaohui Zou, and Ahmedin Jemal.
Cancer statistics, 2014.
CA: A Cancer Journal for Clinicians, 64(1):9–29, 2014.

[5] Eric R Fearon and Bert Vogelstein.
A genetic model for colorectal tumorigenesis.
Cell, 61(5):759–67, Jun 1 1990.

[6] T Jess, C Rungoe, and L Peyrin-Biroulet.
Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies.
Clin Gastroenterol Hepatol., 10(6):639–45, Jun 2012.

[7] Ying Jiang, Qiwen Ben, Hong Shen, Weiqi Lu, Yong Zhang, and Jun Zhu.
Diabetes mellitus and incidence and mortality of colorectal cancer: a systematic review and meta-analysis of cohort studies.

European Journal of Epidemiology, 26(11):863–876, 2011.

[8] AT Chan and EL Giovannucci.
Primary prevention of colorectal cancer.
Gastroenterology., 138((6):):2029–2043, Jun; 2010.

[9] Aleksandar D. Kostic, Dirk Gevers, Chandra Sekhar Pedamallu, Monia Michaud, Fujiko Duke, Ashlee M. Earl, Akinyemi I. Ojesina, Joonil Jung, Adam J. Bass, Josep Tabernero, Jose Baselga, Chen Liu, Ramesh A. Shivdasani, Shuji Ogino, Bruce W. Birren, Curtis Huttenhower, Wendy S. Garrett, and Matthew Meyerson.
Genomic analysis identifies association of fusobacterium with colorectal carcinoma.
Genome Research, 22(2):292–298, 2012.

[10] MR Rubinstein, Wang X, Liu W, Hao Y, Cai G, and Han YW.
Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating e-cadherin/β-catenin signaling via its fada adhesin.
Cell Host Microbe, 14(2):195–206, Aug 14 2013.

[11] G Binefa, Rodriguez-Moranta F, Teule A, and Medina-Hayas M.
Colorectal cancer: from prevention to personalized medicine.
World J Gastroenterol., 20(22):6786–808, Jun 14 2014.

[12] Paul Lichtenstein, Niels V. Holm, Pia K. Verkasalo, Anastasia Iliadou, Jaakko Kaprio, Markku Koskenvuo, Eero Pukkala, Axel Skytthe, and Kari Hemminki.
Environmental and heritable factors in the causation of cancer: Analyses of cohorts of twins from sweden, denmark, and finland.
New England Journal of Medicine, 343(2):78–85, 2000.
PMID: 10891514.

[13] C Lengauer, K. W. Kinzler, and B Vogelstein.
Genetic instability in colorectal cancers.
Nature, 386:623 – 627, April 10 1997.

[14] M Hermsen, C Postma, J Baak, M Weiss, A Rapallo, A Sciutto, G Roemen, JW Arends, R Williams, W Giaretti, A De Goeij, and Meijer G.
Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability.
Gastroenterology, 123(4):1109–19, Oct 2002.

[15] A Tenesa, SM Farrington, Prendergast JG, Porteous ME, M Walker, N Haq, RA Barnetson, E Theodoratou, N Cetnarskyj, Rand Cartwright, C Semple, AJ Clark, FJ Reid, LA Smith, K Kavoussanakis, T Koessler, PD Pharoah, S Buch, C Schafmayer, J Tepel,

S Schreiber, H Volzke, CO Schmidt, J Hampe, J Chang-Claude, M Hoffmeister, H Brenner, S Wilkening, F Canzian, G Capella, V Moreno, IJ Deary, JM Starr, IP Tomlinson, Z Kemp, K Howarth, L Carvajal-Carmona, E Webb, P Broderick, J Vijayakrishnan, RS Houlston, G Rennert, D Ballinger, L Rozek, SB Gruber, K Matsuda, T Kidokoro, Y Nakamura, BW Zanke, CM Greenwood, J Rangrej, R Kustra, A Montpetit, TJ Hudson, S Gallinger, H Campbell, and MG. Dunlop.

Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21.

Nat Genet., 40(5):631–7, May; 2008.

Epub 2008 Mar 30.

[16] Eleanor J. Douglas, Heike Fiegler, Andrew Rowan, Sarah Halford, David C. Bicknell, Walter Bodmer, Ian P. M. Tomlinson, and Nigel P. Carter.

Array comparative genomic hybridization analysis of colorectal cancer cell lines and primary carcinomas.

Cancer Research, 64(14):4817–4825, 2004.

[17] Kentaro Nakao, Kshama R. Mehta, Jane Fridlyand, Dan H. Moore, Ajay N. Jain, Amalia Lafuente, John W. Wiencke, Jonathan P. Terdiman, and Frederic M. Waldman.

High-resolution analysis of dna copy number alterations in colorectal cancer by array-based comparative genomic hybridization.

Carcinogenesis, 25(8):1345–1357, 2004.

[18] C Therkildsen, Jonsson G, Dominguez-Valentin M, Nissen A, Rambech E, Halvarsson B, Bernstein I, Borg K, and Nilbert M.

Gain of chromosomal region 20q and loss of 18 discriminates between lynch syndrome and familial colorectal cancer.

Eur J Cancer., 49(6):1226–35, Apr 2013.

[19] JR Pollack, Perou CM, AA Alizadeh, MB Eisen, A Pergamenschikov, CF Williams, SS Jeffrey, D Botstein, and PO. Brown.

Genome-wide analysis of dna copy-number changes using cdna microarrays.

Nat Genet., 23(1):41–6, Sep 1999.

[20] Paul J. Hurd and Christopher J. Nelson.

Advantages of next-generation sequencing versus the microarray in epigenetic research.

Briefings in Functional Genomics & Proteomics, 8(3):174–183, 2009.

[21] A Kiezun, K Garimella, R Do, NO Stitziel, BM Neale, PJ McLaren, N Gupta, P Sklar, PF Sullivan, JL Moran, CM Hultman, P Lichtenstein, P Magnusson, T Lehner, YY Shugart, AL Price, PI de Bakker, SM Purcell, and SR. Sunyaev.

Exome sequencing and the genetic basis of complex traits.
Nat Genet., 44(6):623–30, May 29 2012.

[22] Chee-Seng Ku, Nasheen Naidoo, and Yudi Pawitan.
Revisiting mendelian disorders through exome sequencing.
Human Genetics, 129(4):351–370, 2011.

[23] Jamie K. Teer and James C. Mullikin.
Exome sequencing: the sweet spot before whole genomes.
Human Molecular Genetics, 19(R2):R145–R151, 2010.

[24] M Zhao, Q Wang, Q Wang, P Jia, and Z. Zhao.
Computational tools for copy number variation (cnv) detection using next-generation sequencing data: features and perspectives.
BMC Bioinformatics., 14, 2013.

[25] Jordi Camps, Quang Tri Nguyen, Hesed M. Padilla-Nash, Turid Knutsen, Nicole E. McNeil, Danny Wangsa, Amanda B. Hummon, Marian Grade, Thomas Ried, and Michael J. Difilippantonio.
Integrative genomics reveals mechanisms of copy number alterations responsible for transcriptional deregulation in colorectal cancer.
Genes, Chromosomes and Cancer, 48(11):1002–1017, 2009.

[26] Shigekazu Hidaka, Toru Yasutake, Hiroaki Takeshita, Masamichi Kondo, Takashi Tsuji, Atsushi Nanashima, Terumitu Sawai, Hiroyuki Yamaguchi, Tohru Nakagoe, Hiroyoshi Ayabe, and Yutaka Tagawa.
Differences in 20q13.2 copy number between colorectal cancers with and without liver metastasis.
Clinical Cancer Research, 6(7):2712–2717, 2000.

[27] B Carvalho, C Postma, S Mongera, E Hopmans, S Diskin, M A van de Wiel, W van Criekinge, O Thas, A Matthai, M A Cuesta, J S Terhaar sive Droste, M Craanen, E Schrock, B Ylstra, and G A Meijer.
Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression.
Gut, 58(1):79–89, 2009.

[28] Yuval Tabach, Ira Kogan-Sakin, Yosef Buganim, Hilla Solomon, Naomi Goldfinger, Randi Hovland, Xi-Song Ke, Anne M. Oyan, Karl-H. Kalland, Varda Rotter, and Eytan Domany.
Amplification of the 20q chromosomal arm occurs early in tumorigenic transformation and may initiate cancer.
PLoS ONE, 6(1):e14632, 01 2011.

[29] A Kallioniemi, OP Kallioniemi, G Citro, Sauter G, S DeVries, R Kerschmann, P Caroll, and F. Waldman.
Identification of gains and losses of dna sequences in primary bladder cancer by comparative genomic hybridization.
Genes Chromosomes Cancer., 12(3):213–9, Mar 1995.

[30] Ben Langmeadand Steven L Salzberg.
Fast gapped-read alignment with bowtie 2.
Nature Methods, 9:357,Äì359, March 2012.

[31] Junbo Duan, Ji-Gang Zhang, Hong-Wen Deng, and Yu-Ping Wang.
Comparative studies of copy number variation detection methods for next-generation sequencing technologies.
PLoS ONE, 8(3):e59128, 03 2013.

[32] Shu Mei Teo, Yudi Pawitan, Chee Seng Ku, Kee Seng Chia, and Agus Salim.
Statistical challenges associated with detecting copy number variations with next-generation sequencing.
Bioinformatics, 28(21):2711–2718, 2012.

[33] Daniel C. Koboldt, Qunyuan Zhang, David E. Larson, Dong Shen, Michael D. McLellan, Ling Lin, Christopher A. Miller, Elaine R. Mardis, Li Ding, and Richard K. Wilson.
Varscan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing.
Genome Research, 22(3):568–576, 2012.

[34] Yuval Benjamini and Terence P. Speed.
Summarizing and correcting the gc content bias in high-throughput sequencing.
Nucleic Acids Research, 40(10):e72, 2012.

[35] Roger Pique-Regi, Alejandro Caceres, and Juan Gonzalez.
R-gada: a fast and flexible pipeline for copy number analysis in association studies.
BMC Bioinformatics, 11(1):380, 2010.

[36] Roger Pique-Regi, Jordi Monso-Varona, Antonio Ortega, Robert C. Seeger, Timothy J. Triche, and Shahab Asgharzadeh.
Sparse representation and bayesian detection of genome copy number alterations from microarray data.
Bioinformatics, 24(3):309–318, 2008.

[37] Hiroshi Iwabuchi, Masaru Sakamoto, Hotaka Sakunaga, Yen-Ying Ma, Maria L. Carcangiu, Daniel Pinkel, Teresa L. Yang-Feng, and Joe W. Gray.

Genetic analysis of benign, low-grade, and high-grade ovarian tumors.
Cancer Research, 55(24):6172–6180, 1995.

[38] Shinichi Fukushige, Frederic M. Waldman, Mitsuhiro Kimura, Tadayoshi Abe, Toru Furukawa, Makoto Sunamura, Masao Kobari, and Akira Horii.
Frequent gain of copy number on the long arm of chromosome 20 in human pancreatic adenocarcinoma.
Genes, Chromosomes and Cancer, 19(3):161–169, 1997.

[39] W. Michael Korn, Toru Yasutake, Wen-Lin Kuo, Robert S. Warren, Colin Collins, Masao Tomita, Joe Gray, and Frederic M. Waldman.
Chromosome arm 20q gains and other genomic alterations in colorectal cancer metastatic to liver, as analyzed by comparative genomic hybridization and fluorescence in situ hybridization.
Genes, Chromosomes and Cancer, 25(2):82–90, 1999.

[40] Rebecca J. Leary, Jimmy C. Lin, Jordan Cummins, Simina Boca, Laura D. Wood, D. Williams Parsons, Sian Jones, Tobias Sjoblom, Ben-Ho Park, Ramon Parsons, Joseph Willis, Dawn Dawson, James K. V. Willson, Tatiana Nikolskaya, Yuri Nikolsky, Levy Kopelovich, Nick Papadopoulos, Len A. Pennacchio, Tian-Li Wang, Sanford D. Markowitz, Giovanni Parmigiani, Kenneth W. Kinzler, Bert Vogelstein, and Victor E. Velculescu.
Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers.
Proceedings of the National Academy of Sciences, 105(42):16224–16229, 2008.

[41] Anke H Sillars-Hardebol, Beatriz Carvalho, Marianne Tijssen, Jeroen A M Belien, Meike de Wit, Pien M Delis-van Diemen, Fredrik Ponten, Mark A van de Wiel, Remond J A Fijneman, and Gerrit A Meijer.
Tpx2 and aurka promote 20q amplicon-driven colorectal adenoma to carcinoma progression.
Gut, 61(11):1568–1575, 2012.

[42] The Cancer Genome Atlas Network.
Comprehensive molecular characterization of human colon and rectal cancer.
Nature, 487(7407):330–7, Jul 18 2012.

[43] H Zhou, J Kuang, L Zhong, WL Kuo, JW Gray, A Sahin, BR Brinkley, and S Sen.
Tumour amplified kinase stk15/btak induces centrosome amplification, aneuploidy and transformation.
Nat Genet., 20(2):189–93, Oct 1998.

[44] YY Yam, B.P. Hoh, S. Othman, N.H.and Hassan, M.M. Yahya, Z. Zakaria, and Ankathil R.

Somatic copy-neutral loss of heterozygosity and copy number abnormalities in malaysian sporadic colorectal carcinoma patients.
Genet. Mol. Res., 12(1):319 – 327, Feb. 2013.

[45] JC Alers, J Rochat, PJ Krijtenburg, WC Hop, R Kranse, C Rosenberg, HJ Tanke, FH Schroder, and H. van Dekken.
Identification of genetic markers for prostatic cancer progression.
Lab Invest., 80(6):931–42, Jun 2000.

[46] Naoto Kuroda, Masato Tamura, Tomoyuki Shiotsu, Shoichiro Nakamura, Takahiro Taguchi, Akira Tominaga, Ondrej Hes, Michal Michal, Chiaki Kawada, Taro Shuin, and Gang-Hong Lee.
Chromosomal abnormalities of clear cell renal cell carcinoma: Frequent gain of chromosome 7.
Pathology International, 60(1):9–13, 2010.

[47] Jin Jen, Hoguen Kim, Steven Piantadosi, Zong-Fan Liu, Roy C. Levitt, Pertti Sistonen, Kenneth W. Kinzler, Bert Vogelstein, and Stanley R. Hamilton.
Allelic loss of chromosome 18q and prognosis in colorectal cancer.
New England Journal of Medicine, 331(4):213–221, 1994.
PMID: 8015568.

[48] Kathleen R. Cho, Jonathan D. Oliner, Jonathan W. Simons, Lora Hedrick, Eric R. Fearon, Antonette C. Preisinger, Philip Hedge, Gary A. Silverman, and Bert Vogelstein.
The {DCC} gene: Structural analysis and mutations in colorectal carcinomas.
Genomics, 19(3):525 – 531, 1994.

[49] Y et al. Takagi.
Somatic alterations of the dpc4 gene in human colorectal cancers in vivo.
Gastroenterology, 111(Issue 5):1369 – 1372, November 1996.

[50] Do Youn et al. Park.
The cables gene on chromosome 18q is silenced by promoter hypermethylation and allelic loss in human colorectal cancer.
The American Journal of Pathology, Volume 171(Issue 5):1509 – 1519, November 2007.

[51] Christina Therkildsen, G. Jansson, Mev Dominguez-Valentin, Anja Nissen, Eva Rambech, Britta Halvarsson, Inge Bernstein, Borg, and Mef Nilbert.
Gain of chromosomal region 20q and loss of 18 discriminates between lynch syndrome and familial colorectal cancer.
European Journal of Cancer, 49(6):1226 – 1235, 2013.

[52] Bert Vogelstein, Eric R. Fearon, Stanley R. Hamilton, Scott E. Kern, Ann C. Preisinger, Mark Leppert, Alida M.M. Smits, and Johannes L. Bos.
Genetic alterations during colorectal-tumor development.
New England Journal of Medicine, 319(9):525–532, 1988.
PMID: 2841597.

[53] SJ Baker, ER Fearon, JM Nigro, SR Hamilton, Preisinger, and B. Vogelstein.
Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas.
Science, 244(4901):217–21, Apr 14 1989.

[54] Alexander T. El Gammal, Michael Bruchmann, Jozef Zustin, Hendrik Isbarn, Olaf J.C. Hellwinkel, Jens Kollermann, Guido Sauter, Ronald Simon, Waldemar Wilczak, Jorg Schwarz, Carsten Bokemeyer, Tim H. Brummendorf, Jakob R. Izbicki, Emre Yekebas, Margit Fisch, Hartwig Huland, Markus Graefen, and Thorsten Schlomm.
Chromosome 8p deletions and 8q gains are associated with tumor progression and poor prognosis in prostate cancer.
Clinical Cancer Research, 16(1):56–64, 2010.

[55] Valentina Boeva, Tatiana Popova, Kevin Bleakley, Pierre Chiche, Julie Cappo, Gudrun Schleiermacher, Isabelle Janoueix-Lerosey, Olivier Delattre, and Emmanuel Barillot.
Control-freec: a tool for assessing copy number and allelic content using next-generation sequencing data.
Bioinformatics, 28(3):423–425, 2012.

[56] Jarupon Fah Sathirapongsasuti, Hane Lee, Basil A. J. Horst, Georg Brunner, Alistair J. Cochran, Scott Binder, John Quackenbush, and Stanley F. Nelson.
Exome sequencing-based copy-number variation and loss of heterozygosity detection: Exomecnv.
Bioinformatics, 27(19):2648–2654, 2011.

[57] Vincent Plagnol, James Curtis, Michael Epstein, Kin Y. Mok, Emma Stebbings, Sofia Grigoriadou, Nicholas W. Wood, Sophie Hambleton, Siobhan O. Burns, Adrian J. Thrasher, Dinakantha Kumararatne, Rainer Doffinger, and Sergey Nejentsev.
A robust model for read count data in exome sequencing experiments and implications for copy number variant calling.
Bioinformatics, 28(21):2747–2754, 2012.

[58] Jason Li, Richard Lupat, Kaushalya C. Amarasinghe, Ella R. Thompson, Maria A. Doyle, Georgina L. Ryland, Richard W. Tothill, Saman K. Halgamuge, Ian G. Campbell, and Kylie L. Gorringe.
Contra: copy number analysis for targeted resequencing.
Bioinformatics, 28(10):1307–1313, 2012.

[59] Christopher A. Miller, Oliver Hampton, Cristian Coarfa, and Aleksandar Milosavljevic.
Readdepth: A parallel r package for detecting copy number alterations from short sequencing reads.
PLoS ONE, 6(1):e16327, 01 2011.

[60] Xutao Deng.
Seqgene: a comprehensive software solution for mining exome- and transcriptome- sequencing data.
BMC Bioinformatics, 12(1):267, 2011.

A

Supplementary Tables

CRC patients (n=42)	
Gender	
Male	31 (73.8%)
Female	11 (26.2%)
Median age (range,years)	70 (43-84)
Site	
Right	12 (28.6%)
Left	30 (71.4%)
Stage	
II A	38 (90.5%)
II B	4 (9.5%)
Recurrence	
No relapse	21 (50%)
Relapse	21 (50%)
Recurrence-free median time (range,months)	60.9 (6.8-127.8)

Table A.1: **Characteristics of the patients included in the study.**

SNPs	
SNP	Position
rs6983267	8q24
rs4939827	18q21
rs16892766	8q23.3
rs10795668	10p14
rs9929218	16q22.1
rs961253	20p12.3
rs11169552	12q13.12
rs4444235	14q22.2
rs10411210	19q13.11
rs3802842	11q23.1
rs6691170	1q41
rs10936599	3q26.2
rs4925386	20q13.33

Table A.2: **List of SNPs used to perform the dynamic arrays in order to ensure that normal and tumor tissues were paired.**

Tool	Language	Input	comments	Ref
Control - FREEC	C++	SAM/BAM/pileup/Eland, BED, SOAP, arachne, psi (BLAT) and Bowtie	No need of controls; normalizes with GC content. Takes into account polyploidy	55
VarScan2	Java	BAM/Pileup	Minimum coverage 20X. Matched tumor-normal samples as input. Depth of high quality bases individually for tumor and normal samples. Needs extra tools for segmentation (CBS suggested)	33
ExomeCNV	R	BAM	Assumes Gaussian distribution to characterize the read depth from WES. Requires matched case-control samples as input. Models the log-ratios using a Geary-Hinkly transformation	56
ExomeDepth	R	BAM	Uses beta-binomial model to fit read depth of WES data. Multiple samples as imput. Is more conservative than exomeCNV.	57
Contra	Python	SAM/BAM	Coverage profiles of the samples calculated with BEDTools. Adjusted coverage to remove GC-bias. Multiple samples or matched case-control. It relies on the case sample being largely copy number neutral.	58
ReadDepth	R	BED	GC correction. Single end or paired end. CBS algorithm, loess regression. Does not require a matched normal sample.	59
SeqGene	Python, R	SAM/pileup	Log2 RPKM estimation of each exon. Reference needed. Not need to be matched. Coverage quantification and visualization	60

Table A.3: **Alternative software that could be used to analyze WES data.** The software used for this study was Varscan2, but instead of using the circular binary segmentation (CBS) as suggested, we performed the segmentation step using R-GADA R package.

B

Supplementary Figures

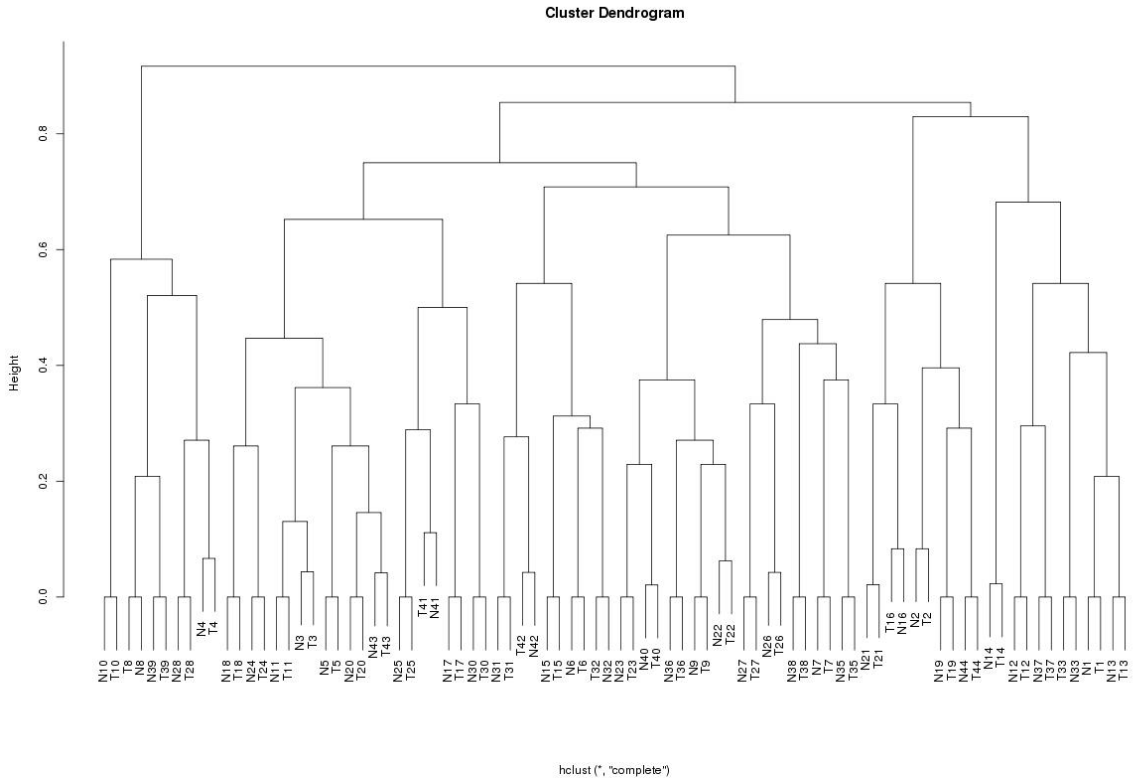

Figure B.1: **Cluster Dendogram of the Normal and Tumor Samples.** It can be observed that all the samples matched their Normal with their Tumoral.

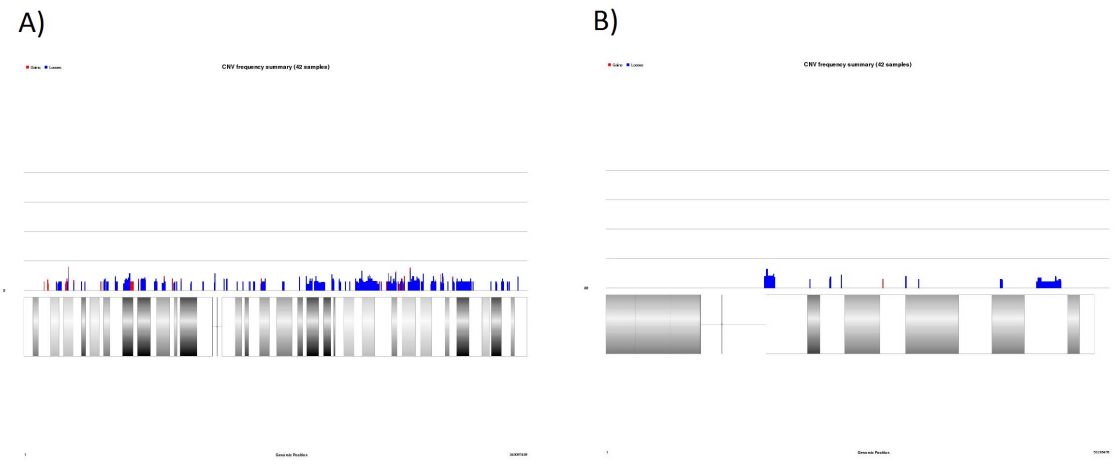

Figure B.2: **Summary of the segments found in chromosome 2(A) and 22(B) across all samples.** Segments are represented in blue for lost and red for gain of DNA. The height of the segment represents the frequency across all samples, being the represented lines a 25%, 50%, 75% and 100% of them. It can be observed that the majority of the segments shown in this cromosome are of small length, in both chromosome 2 and 22.

C

Code

C.1 Copynumber estimation

```
#!/bin/bash

path_data="/mnt/hydra/raid52/exomes/CLX/resultados_intermedios/results/Alineamientos/filtered"

path_results="/mnt/hydra/raid51/ubs/extrahome/43634313R_extra/Dades_exome_seq/results/results_C

samtools="/share/apps/samtools/samtools"
varscan="/share/apps/varscan/VarScan.v2.3.3.jar"

#########################
#
# Run varscan copynumber on normal and tumor mpileup output
#
ref_fasta="/mnt/hydra/raid51/ubs/extrahome/43634313R_extra
/Dades_exome_seq/ucsc.hg19.fasta"

cd $path_data
for fl in *_N.bam; do
out=${fl/_N.bam/}
fl_T=${fl/_N/_T}
$samtools mpileup -q 1 -f $ref_fasta $fl $fl_T

/usr/java/latest/bin/java -jar ${varscan} copynumber --mpileup  ${path_results}/$out.500
--max-segment-size 500 --min-segment-size 100

done
```

```
#########################
#
# Run Varscan copyCaller to adjust for GC content
#

cd ${path_results}

for fl in *500.copynumber; do

out=${fl/.copynumber/}
/usr/java/latest/bin/java -jar ${varscan} copyCaller $fl
--output-file ${path_results}/$out.copynumber.called --recenter-up 0.3

done
```

C.2 Segmentation

```
###############
#
# R_GADA tots
#

mypath <- "/mnt/hydra/raid51/ubs/extrahome/43634313R_extra/Dades_exome_seq/
            results/results_CNV"

library(gada)

copynumbers <- list.files(path = mypath, pattern = "500.copynumber.called",
                          all.files =FALSE,
                          full.names = TRUE,
                          recursive = TRUE,
                          ignore.case = TRUE)

copynumbers<- copynumbers[-(grep("gc",copynumbers,value=F))]

sample <- list.files(path = mypath,
```

```
                    pattern = "500.copynumber.called",
                    all.files = FALSE,
                    full.names = FALSE,
                    recursive = TRUE,
                    ignore.case = TRUE)

sample<- sample[-(grep("gc",sample,value=F))]
sample <- strsplit(x=sample, split=".500.copynumber.called")


for (i in 1:length(copynumbers)){
  j<- sample[i]
  print(j)
        A <- read.table(copynumbers[i], header = T)
        print("taula")
        A$chromosome <- gsub(as.character(A$chrom),
                        pattern = 'chr',
                        replacement = '') ##remove the annoying chr letters

        chrr <- c(1:22, "X", "Y")
  probe = paste("rm",A$chr_stop[A$chromosome %in% chrr], sep='')
        gen.info <- data.frame(probe = probe,
                chr = A$chromosome[A$chromosome %in% chrr],
                pos= A$chr_start[A$chromosome %in% chrr])
  #
        print("gen.info")
        mydata <- A$adjusted_log_ratio[A$chromosome %in% chrr]
        gadadata <-setupGADAgeneral(mydata,gen.info=gen.info)

        print("SBL")
        step1<-SBL(gadadata, estim.sigma2=TRUE, aAlpha=0.8)

        print("BackwardElimination")
        step2<-BackwardElimination(step1, T=5, MinSegLen=3)

        a <- summary(step2)
        a <- a[-(grep("0", a$State, value=F)),]
        print(paste(j, "result.csv", sep="_"))
```

```
        write.csv(a, file=(paste(j, "result.csv", sep="_")))

        print(paste(j, "plots.eps", sep="_"))
        postscript(file=(paste(j, "plots.eps", sep="_")))

        for (k in 1:(length(chrr)-2)){
        plotRatio(step2, chr=k, postscript=TRUE, segments=TRUE)
        }
        plotRatio(step2, chr="X", segments=TRUE, postscript=TRUE)
        plotRatio(step2, chr="Y", segments=TRUE, postscript=TRUE)
        dev.off()

        print('Next')
        }


################################
```

C.3 Summary of the recurrent CNV

```
path <- "/mnt/hydra/raid51/ubs/extrahome/43634313R_extra/Dades_exome_seq/results/results_CNV"

results <- list.files(path = path,
                      pattern = "_result.csv",
                      all.files =FALSE,
                      full.names = TRUE,
                      recursive = TRUE,
                      ignore.case = TRUE)

results_noms <-list.files(path = path,
                          pattern = "_result.csv",
                          all.files =FALSE,
                          full.names = FALSE,
                          recursive = TRUE,
                          ignore.case = TRUE)
```

```
results_noms <- as.character(strsplit(x=results_noms, split="_result.csv"))

b <- NULL

for (p in 1:length(results)){
  a <- read.csv(results[p], header=T)
a$Sample <- p
b <- rbind(b,a)
}

tail(b)

######################
#
# Load at least 1 gen.info
#

library(gada)
A2027 <- read.table("/results/A2027.copynumber.called.recentered", header = T)

# remove the annoying chr letters
A2027$chromosome <- gsub(as.character(A2027$chrom),
                         pattern = 'chr',
                         replacement = '')

chrr <- c(1:22, "X", "Y")

probe <- paste("rm",A2027$chr_stop[A2027$chromosome %in% chrr], sep='')

gen.info <- data.frame(probe = probe,
                       chr = A2027$chromosome[A2027$chromosome %in% chrr],
                       pos= A2027$chr_stop[A2027$chromosome %in% chrr])

######################
#
#
# Load functions
#
```

```
#

######################
# drawChromosome as in R-GADA package
######################

drawChromosome<- function (chr, size = 0.05, limits, print.names = TRUE, ...)
{
    on.exit(par(xpd = TRUE))
    par(xpd = FALSE)
    if (chr == 23)
        chr <- "X"
    if (chr == 24)
        chr <- "Y"
    chromName <- paste("chr", chr, sep = "")
    if (missing(limits))
        limits <- c(0, tamanysChroms$size[tamanysChroms$chrom ==
            chromName])
    plot(0, axes = FALSE, xlab = "", type = "n", col = "gray",
        xlim = limits, ...)
    pp <- par("usr")
    ytop <- pp[3]
    ybottom <- pp[3] + size
    ymig <- (ytop + ybottom)/2
    segments(0, ymig, tamanysChroms[tamanysChroms$chrom == chromName,
        2], ymig, col = "gray", cex = 2)
    bandes <- ideogram[ideogram$chrom == chromName, ]
    bandes$posGrafic <- bandes$chromStart + (bandes$chromEnd -
        bandes$chromStart)/2
    numBandes <- dim(bandes)[1]
    gieStain <- levels(bandes$gieStain)
    colorsBandes <- c("red", "gray100", "black", "gray25", "gray50",
        "gray75", "gray50", "gray50")
    japintat <- 0
    for (h in 1:numBandes) {
        chromStart <- bandes$chromStart[h]
        chromEnd <- bandes$chromEnd[h]
        giemsaBand <- bandes$gieStain[h]
```

```
        colorBanda <- colorsBandes[giemsaBand]
        if (giemsaBand == "acen") {
            segments(chromStart, ybottom, chromStart, ytop, col = "black")
        }
        else {
            if (giemsaBand == "gvar") {
                rect(chromStart, ybottom, chromEnd, ytop, col = colorBanda,
                  angle = 90, density = 0.9)
                cylindrect(chromStart, ybottom, chromEnd, ytop,
                  col = colorBanda, gradient = "y", nslices = 50)
            }
            else {
                rect(chromStart, ybottom, chromEnd, ytop, col = colorBanda,
                  border = "gray50")
                cylindrect(chromStart, ybottom, chromEnd, ytop,
                  col = colorBanda, gradient = "y", nslices = 50)
            }
        }
        if (japintat == 0 & giemsaBand == "acen") {
            japintat <- 1
            points(chromEnd, ymig, pch = 19, col = "gray", xpd = TRUE,
                cex = 0.8)
        }
    }
    if (print.names)
        text(bandes$posGrafic, rep(ybottom - 1.5 * size, numBandes),
            bandes$name, srt = -90, adj = 0, cex = 0.6, xpd = TRUE)
}


##############################################################
################## Count.altered #############################
##############################################################

count.altered <- function(i, x){
final.df <- NULL
for (k in 1:length(i)){
segments <- nrow(x[x$chromosome==k,])
```

```
gains_df <- x[,7] == 1
Gains <- nrow(x[gains_df & x$chromosome == k,])
loss <- x[,7] == "-1"
Losses <- nrow(x[loss & x$chromosome == k,])

chromosome <- paste("chr",k,sep=' ')
df.chr <- data.frame(
chromosome,
segments,
Gains,
Losses
)

final.df <- rbind(final.df, df.chr)
}
Freq.g <- NULL
Freq.l <- NULL
Segments <- nrow(x)
NSamples <- max(x$Sample)
Gains <- nrow(x[x$State == 1,])
Losses <- nrow(x[x$State == "-1",])
options(digits=4)
Per.Gains <- (Gains/Segments) * 100
options(digits=4)
Per.Losses <- (Losses/Segments) * 100
info <- data.frame(
Segments,
Gains,
Per.Gains,
Losses,
Per.Losses)
Min <- min(x$LenProbe)
First_Qu <- quantile(x$LenProbe, 0.25, type=7)
Median <- median(x$LenProbe)
Mean <- mean(x$LenProbe)
Third_Qu <- quantile(x$LenProbe, 0.75, type=7)
Max <- max(x$LenProbe)
info2 <- data.frame(
```

```
Min,
First_Qu,
Median,
Mean,
Third_Qu,
Max)
i.bli <- list("Total_Segments" = info,
                "Summary_segments" = info2,
                "Altered" = final.df)
return(i.bli)
}

#################################################################
######### My function to properties segments ################
#################################################################

def.segments <- function(x, chr){

lost <- x[x$State == "-1",]
lost <- lost[lost$chromosome == chr,]
won <- x[x$State == 1,]
won <- won[won$chromosome == chr,]
start.l <- lost[,2]
end.l <- lost[,3]
list.a <- NULL
losses <- NULL

if (nrow(lost)!=0){
for (i in 1:nrow(lost)){
a <- start.l[i]:end.l[i]
list.a <- c(list.a,a)
}
for (k in 1:nrow(lost)){
a <- start.l[k]:end.l[k]
losses.a <- data.frame(
chr = chr,
pos = lost[k,2],
pos.end = lost[k,3],
```

```
Freq.l = sum(duplicated(list.a[a]))/length(a) + 1
)
losses <- rbind(losses, losses.a)
}
  }
start.g <- won[,2]
end.g <- won[,3]
list.b <- NULL
list.gains <- NULL
gains <- NULL
if (nrow(won) != 0){
for (j in 1:nrow(won)){
b <- start.g[j]:end.g[j]
list.b <- c(list.b,b)
}
for (t in 1:nrow(won)){
b <- start.g[t]:end.g[t]
gains.b <- data.frame(
chr = chr,
pos = won[t,2],
pos.end = won[t,3],
Freq.g = sum(duplicated(list.b[b]))/length(a) + 1
)
gains <- rbind(gains, gains.b)
}
}

bla <- list("gains"=gains, "losses" = losses)
return(bla)

}

#################################################################
######################### my_plotWG #############################
#################################################################

my_plotWG <- function(x, min.percentage = 0.03,
                      max.number.cnv = 400,
```

```
                      length.base = c(10,10^20),
                      chr)
{
    if (missing(chr))
  chr = c(1:22, "X", "Y")
    gen.info <- gen.info[gen.info$chr %in% chr,]
    require(gada)
    require(plotrix)
    data(genomicInfo)
    if (!"gen.info" %in% ls(pos = .GlobalEnv)) {
        stop("gen.info of at least one sample is required in '.GlobalEnv'")
    }
    col.legend <- c("red", "blue")
    chrs <- gen.info[, 2]
    limits <- c(min(gen.info[, 3]), max(gen.info[, 3]))
    n.chr <- length(chr)
    old.mfrow <- par("mfrow")
    old.mar <- par("mar")
    old.xpd <- par("xpd")
    on.exit(par(mfrow = old.mfrow, mar = old.mar))
    par(mfrow = c(n.chr + 2, 1))
    par(mar = c(0.5, 5, 0.2, 3))
    par(xpd = TRUE)
    plot(c(1:3), rep(1, 2, 3), axes = FALSE, xlab = "", ylab = "",
        type = "n")
    legend(1, 1, c("Gains", "Losses"), col = col.legend, pt.bg = col.legend,
        pch = rep(22, 4), horiz = TRUE, cex = 1, bty = "n", pt.cex = 1.6,
        yjust = 0.5)
    nSamples <- max(x$Sample)  #attr(x, "Samples")[2]
    text(2, 1, paste("CNV frequency summary (", nSamples, " samples)",
        sep = ""), font = 2)
    for (i in 1:length(levels(gen.info$chr))) {
        select <- gen.info[, 2] == i
        pos.chr <- gen.info[select, 3]
        drawChromosome(i, 0.5, print.names = FALSE, ylim = c(-0.5,
            1), limits = limits, ylab = "")
        ss <- c(0, 0.25, 0.5, 0.75, 1)
        for (j in 1:5) {
```

```
            segments(limits[1], ss[j], max(pos.chr), ss[j], col = "gray60")
        }
print(paste("def.segments of chr", i, sep=' '))
        altered <-  def.segments(x, chr = i)
print(paste("drawing segments of chr", i, sep=' '))
print(paste("chr", i, "has", length(altered$gains$pos), "gain segments", sep=' '))

        if (!is.null(altered$gains)) {
            gains <- altered$gains
            pos.gains <- as.numeric(gains$pos)
    pos.end <- as.numeric(gains$pos.end)
            hh2 <- as.numeric(gains$Freq.g)/nSamples
            hh2[hh2 < min.percentage] <- NA
            if (length(pos.gains) > 0)
for (j in 1:length(pos.gains)){
  print(paste("segment", j, sep=' '))
      segments(pos.gains[j]:pos.end[j], 0,
               pos.gains[j]:pos.end[j], hh2[j],
               col = col.legend[1])
}
        }
print(paste("chr", i, "has", length(altered$losses$pos), "lost segments", sep=' '))
        if (!is.null(altered$losses)) {
            losses <- altered$losses
            pos.losses <- as.numeric(losses$pos)
            pos.end.l <- as.numeric(losses$pos.end)
            hh <- as.numeric(losses$Freq.l)/nSamples
            hh[hh < min.percentage] <- NA
            if (length(pos.losses) > 0) {
        for (k in 1:length(pos.losses)){
        print(paste("segment", k, sep=' '))
            segments(pos.losses[k]:pos.end.l[k], 0,
                     pos.losses[k]:pos.end.l[k], hh[k],
                     col = col.legend[2])
}
        }
        text(par("usr")[1], 0, chr[i], cex = 1, adj = 0.5)
        text(par("usr")[1], 0, chr[i], cex = 1, adj = 0.5)
```

```
    }
    plot(limits, c(1, 1), axes = FALSE, xlab = "", ylab = "",
        type = "n", xlim = limits)
    text(limits[1], 1, limits[1], cex = 1, adj = 0)
    text(limits[2], 1, limits[2], cex = 1, adj = 1)
    text((limits[2] - limits[1])/2, 1, "Genomic Position", cex = 1,
        adj = 0, font = 2)
}
```

D

Genes

D.1 Genes present in Lost segments

TPI1P1 ANKRD13C CNTN5 OR5T3 SNORD113-3 SNORD114-12 HS3ST3A1 CENPV NDEL1
MSH4 ASB17 RN7SL488P ARHGAP32 SNORD113-4 SNORD114-13 NCOR1 MYH4 RPS27AP1 PITPNM3
ST6GALNAC3 ITGB3BP OR4C50P SNORD113-5 SNORD114-22 MYH1 MYH2 MAGOH2 MTND1P14
SDHDP6 EFCAB7 ENDOD1 SNORD113-6 SNORD114-23 RPL19P18 LINC00675 TRNAQ41P
C1orf63 RHD RPL5P6 MIR4490 SNORD113-7 SNORD114-24 NTN1 STX8 RNU6-1065P MTND2P12
SYCP1 TSHB ZNF644 NAALAD2 SNORD113-8 SNORD114-25 WDR16 USP43 RPL15P21 SLC47A1
C1orf173 FPGT CASP3P1 CHORDC1 SNORD113-9 SNORD114-26 DHRS7C RN7SL601P FAM106A
FPGT-TNNI3K PTGER3 RNA5SP345 SNORD114-1 SNORD114-27 MYH13 MYH8 MYH3 SCO1 TRIM16L
TNNI3K PRKACB CEP57 SNORD114-14 SNORD114-28 RPL21P122 ADPRM FBXW10
TSPAN2 FEN1P1 MTMR2 SNORD114-15 SNORD114-29 LINC00670 TMEM220 TVP23B
LRRIQ3 CRYZ CDC7 WDR36 FAM111B SNORD114-16 SNORD114-30 ZNF18 PIRT SHISA6 PRPSAP2
HFM1 LRRC7 RICTOR FAM111A SNORD114-17 OR4K15 MYOCD RPL23AP76 NOS2P2
TXNP2 DE- SRFBP1 CBX3P7 GRM5 SNORD114-18 OR4K16P DNAH9 CDRT7 USP32P2
PDC1 C1orf168 CSNK1G3 TYR WDR36 SNORD114-19 OR4N1P MIR1269B CDRT8 PMP22 SRP68P2
KIAA1107 OSMR RICTOR SNORD114-20 OR4K6P RN7SL550P PIK3R6 CCDC144B
WDR3 SPAG17 SLC25A46 SRFBP1 SNORD114-21 OR4K3 NGDN COX10-AS1 CCDC42 SP- RN7SL639P
RN7SKP285 ADAMTS19 CSNK1G3 MIPOL1 TC2N OR4N5 ARHGAP44 DYE4 MFSD6L FOXO3B
SOD2P1 DMXL1 OSMR GMFB CEP128 WDR36 ELAC2 PIK3R5 UBE2SP2
COL11A1 SLC12A2 SLC25A46 OR4Q2 RICTOR SPATA22 KIAA0753 RN7SL627P
RNPC3 TYW3 PGGT1B ADAMTS19 OR4K14 SRFBP1 NPM1P45 TXNDC17 RN7SL775P
RN7SL538P CCDC112 DMXL1 GPHN OR4N2 CSNK1G3 GLP2R MED31 MPRIP VPS53
PIN1P1 RPE65 HECTD2 SLC12A2 OR4Q3 OR4M1 OSMR RCVRN GAS7 RNA5SP435 RPS4XP17
NEGR1-IT1 TM9SF3 CC- PGGT1B OR11K2P SLC25A46 TBC1D28 MIR4731 XAF1 FBXO39
GDI2P2 SER2 EXOC6 CCDC112 OR4H12P ADAMTS19 ZNF286B RNU6-799P TEKT1 PEMT
RPL31P12 KIF20B LRRK2 OR4K2 DMXL1 RN7SL792P RPL9P2 SNORD91B
RNU6-1246P ZNF518A PARPBP OR4K5 OR4K1 SLC12A2 TEKT3 CDRT4 ZNF29P SNORD91A
KRT8P21 WDR36 PMCH WDR36 OR4K4P PGGT1B TVP23C- UBE2SP1 SRR TSR1
RN7SKP19 RICTOR RICTOR OR4K13 CCDC112 CDRT4 CDRT15P2 C17orf85
RNA5SP50 SRFBP1 SRFBP1 OR4L1 SLC24A5 TVP23C ZSWIM5P1 ULK2 AKAP10
RNU4ATAC8P CSNK1G3 CSNK1G3 OR4K17 MYEF2 RNU6- MYH10 RNA5SP436 SPECC1
NEGR1 OSMR OSMR OR4U1P 449P WDR72 RN7SL129P IL6STP1 KCTD9P1
LRRC53 SLC25A46 SLC25A46 RNA5SP380 UNC13C ALDH3A2 MEIS3P1 RNU6-258P
ZRANB2-AS1 ADAMTS19 ADAMTS19 OR4T1P HNRNPA1P74 RPS18P12 FTLP12 DHRS7B
MIR186 DMXL1 DMXL1 ARHGAP5 RSL24D1 COX10 FAM211A- OR1E3 OR1P1
ZRANB2-AS2 SLC12A2 SLC12A2 SNORD113-1 RNU2-53P CDRT15 AS1 CDRT1 OR1D4
ZRANB2 PGGT1B PGGT1B SNORD114-2 DYX1C1- HS3ST3B1 TRIM16 OR1D3P
LPHN2 CCDC112 CCDC112 SNORD114-3 CCPG1 CDRT15P1 ZNF286A OR1D5 OR1D2
MAN1A2 NOX4 FAM76B KTN1 RNU6- SNORD114-4 MIR628 RPLP1P11 TBC1D26 OR1G1 OR1A2
VPS13D RNU6- DISC1FP1 1277P TTC6 SNORD114-5 RAB27A RNU6-314P ADORA2B OR1A1 DOC2B
622P LHX8 MIR1261 C14orf39 SNORD114-6 PIGB CCPG1 RNU6-862P ZSWIM7 DNAH2
SLC44A5 OSBPL9P2 RPA2P1 SNORD114-7 MIR744 RN7SL442P UBB TRPV2 ZNF594
RN7SL242P OSBPL9P3 RPL13AP2 SNORD114-9 MAP2K4 MIR1288 FAM211A SCIMP
LRRC40 SESN3 MDGA2 MEG8 SNORD114-10 RNU11-2P TTC19 PIGL KRBA2 RPL26 RABEP1
SRSF11 RPA2P3 SNORD113-2 SNORD114-11 MIR548H3 RNU7-43P RNF222 NUP88 ZZEF1

CYB5D2 SLC47A2 P2RX1 CRK MYO1C TMEM95 CD226 RTTN RNU2-69P MBP GALR1
ANKFY1 ALDH3A1 ATP2A3 INPP5K TNK1 SERPINB13 OACYLP SALL3 ATP9B
UBE2G1 LGALS9B KCNAB3 DBIL5P TMEM256- CPLX4 TXNL1 RPS26P54 NFATC1
RNA5SP434 CDRT15L2 TRAPPC1 MIR3183 PLSCR3 MAP1LC3P GLUD1P4 CTDP1
RYKP1 MIEF2 CCDC144NL CNTROB PITPNA-AS1 SGSM2 RNA5SP459 NFE2L3P1 KCNG2 PQLC1
TOP3A SMCR8 USP22 GUCY2D RN7SL105P TMEM256 MIR4529 RN7SL342P HSBP1L1
SHMT1 TMEM11 ALOX15B MIR22HG NLGN2 RPL21P126 SDCCAG3P1 TXNL4A
EVPLL C17orf103 ALOX12B PITPNA DVL2 SPEM1 RAB27B FAM60CP RBFA ADNP2
LGALS9C MAP2K3 ALOXE3 SLC43A2 C17orf74 CCDC68 TCF4 RNU6-567P PARD6G
NEK4P2 KCNJ12 HES7 PER1 MIR132 TMEM102 RPSAP57 RPS3AP49 TCEB3CL2
RNASEH1P2 C17orf51 SPNS3 VAMP2 MIR212 FGF11 DYNAP RNU4-17P TCEB3CL
RNU6-405P UBBP4 TMEM107 RN7SL624P CHRNB1 CCDC178 MRPS5P4 RNA5SP458
RN7SL620P MTRNR2L1 C17orf59 HNRNPA1P16 ZBTB4 SMAD2 CTBP2P3 RPLP0P11
UPF3AP1 SNORD3B-1 AURKB SPNS2 RN7SL33P SLC35G6 HNRNPA1P11 RNU6-116P HNRNPA3P16
USP32P1 SNORD3B-2 MYBBP1A EIF4A1P9 POLR2A MNT CBLN2 RPL30P14 RPL17P46
SRP68P1 KYNUP2 CTC1 GGT6 SAMD11P1 METTL16 SNRPGP2 RPIAP1 RN7SL695P
FAM106CP SNORD3D SMTNL2 RN7SL608P TNFSF12 C18orf54 RPL17P44 SRSF10P1
KRT16P2 KYNUP1 PFAS ALOX15 PHF23 PAFAH1B1 POLI STARD6 ACTBP9 RNU1-46P
KRT17P1 SNORD3A PELP1 SCARF1 TNFSF12- RPS2P6 RN7SL705P SS18L2P2
COTL1P1 GRAPL SLC25A35 MIR1253 TNFSF13 RN7SL795P RNU6-142P RSL24D1P9
TBC1D27 SNORD3C RANGRF RN7SL605P TNFSF13 RNA5SP460 ATP5G1P6 RPS8P3
MPRIP-AS1 KYNUP3 ARRB2 RILP CLUH SENP3 MIR548AV BOD1L2 MBD1 CXXC1
RNU6-767P EPN2-AS1 ARHGEF15 GABARAP EIF4A1 RN7SL401P ST8SIA3 SKA1 MAPK4
RPL13P12 MIR1180 ODF4 MED11 PRPF8 RAP1GAP2 RN7SL551P ONECUT2 MRO ME2
TSEN15P1 RPS2P46 CXCL16 CTDNEP1 TRPV3 CD68 SOCS6 FECH NARS ELAC1 SMAD4
RNU6-468P RPL17P43 ZMYND15 RN7SL774P MPDU1 GTSCR1 ATP8B1 MEX3C
SMCR2 RAI1- SNORA59B TM4SF5 VMO1 TXNP4 TRPV1 SOX15 NETO1 NEDD4L MIR4527 TP-
AS1 SMCR5 EPN2 GLTPD2 RNU6-955P FXR2 SHPK FBXO15 ALPK2 MALT1 MTP1 TCEB3C
MIR33B RNU6-1057P PSMB6 ATP6V0CP1 CTNS SHBG TIMM21 ZNF532 LINC00470
RPL7AP65 CCDC144CP PLD2 MINK1 RN7SL171P TAX1BP3 CYB5A SEC11C GRP GALNT1
RPL21P121 B9D1 CHRNE RN7SL784P P2RX5- C18orf63 RAX CCBE1 NOL4 DTNA
TBC1D3P4 RN7SL17P C17orf107 ELP5 TLCD2 TAX1BP3 ASXL3 PMAIP1 DSG4 MIB1
YWHAEP2 UPF3AP2 GP1BA RNU7-31P EMC6 P2RX5 RPL29P32 MC4R CDH20 MIR133A1
KRT17P2 USP32P3 SLC25A11 RNU6-1264P SAT2 ATP1B2 SNORA37 RNF152 RNU6-443P
KRT16P1 SRP68P3 RNF167 PFN1 BTF3P14 TP53 ITGAE MBD2 PIAS2 TNFRSF11A DSG1 RNU6-
KRT16P4 FAM106B ENO3 SPAG7 CLDN7 GSG2 WRAP53 LINC00907 ZCCHC2 1131P TCEB3B
TNPO1P2 NOS2P3 CAMTA2 MIR4520B EFNB3 RNA5SP454 PHLPP1 BCL2 RNU6-708P
SLC5A10 YWHAEP3 INCA1 KIF1C RPL23AP73 KDM6B PIK3C3 RIT2 KDSR VPS4B RNU7-191P
FAM83G KRT16P5 SLC52A1 ALOX12P2 TMEM88 SYT4 CDH2 SERPINB5 RNA5SP456
GRAP ZNF287 KRT16P3 WSCD1 AIPL1 MIR497HG LSMD1 NAPG PIEZO2 SERPINB12 MIR4743
ZNF624 TBC1D3P3 FAM64A RPL7AP64 CYB5D1 CCDC58P1 RAB31 RNU6- HDHD2
CCDC144A COTL1P2 RPH3AL MIR324 TMX3 WDR7 PMM2P1 903P KRT18P8 MIR4744
TNFRSF13B ZSWIM5P2 C17orf97 WDR81 KIAA1468 ANKRD12 RN7SL862P MIR1539
PLD6 FLCN MEIS3P2 C17orf100 SLC2A4 LMAN1 NDUFV2 RNA5SP449 SNORD58C
COPS3 NT5M RNFT1P3 FAM101B SENP3- SERPINB11 TWSG1 FAUP1 SNORD58A
MED9 RASD1 HNRNPA1P19 SLC13A5 EIF4A1 SERPINB3 RALBP1 LINC00909 SNORD58B
RAI1 SREBF1 OLA1P2 ALOX12 SNORA67 DCC PIGN RNU2-27P LINC00683 SRP72P1
TOM1L2 SCDP1 RNASEK SNORA48 SERPINB4 AQP4-AS1 ARL2BPP1 SMUG1P1
LRRC48 ABHD17AP6 FAM57A SNORD10 PRPF19P1 UBA52P9 RNU6-346P IER3IP1
ATPAF2 MAPK7 RNASEK- RPL29P2 CDH7 CDH19 RBM22P1 RPL26P35 SCARNA17
GID4 DRG2 RNU6-1178P C17orf49 YBX2 SERPINB2 PA2G4P3 BDP1P RNA5SP457
MYO15A AL- RNASEH1P1 GEMIN4 SCARNA21 SERPINB10 CHST9 RNA5SP461 ADAD1P2
KBH5 LLGL1 RN7SL426P GLOD4 MIR4314 RNU6-1037P PPP4R1 LINC01029 RN7SL310P
FLII ZFP3 PDLIM1P2 C17orf49 SNORD118 MIR5011 KATNAL2 RNU6-655P SKOR2
ZNF232 USP6 RPL21P120 BCL6B LINC00324 RPL31P9 WDR7-OT1 RBFADN ZBTB7C CTIF
RPAIN C1QBP NCOR1P2 SLC16A13 SERPINF2 AKR1B10P2 LINC-ROR SLC25A6P4 SMAD7 DYM
OR3A4P FTLP13 SLC16A11 EIF5A SER- RNU6-39P RNU6-737P FAM69C C18orf32
OR3A5P MFAP4 RNMTL1 PINF1 GPS2 SDHCP1 DSEL RNU6-742P CNDP2 RPL17-
OR1AC1P FAM27L CLEC10A NXN SMYD4 RPA1 CCDC102B RSL24D1P11 CNDP1 C18orf32
OR1R1P RNF112 ASGR2 ASGR1 RTN4RL1 RPL12P39 HMGN1P30 ZNF407 RPL17 LIPG
OR3A2 OR3A1 MTND1P15 TIMM22 DLG4 NEURL4 LINC00305 MIR122 ZADH2 TSHZ1 ACAA2
OR1E1 OR3A3 TRNAS30P ABR ACADVL DPH1 OVCA2 RNU7-146P RPL9P31 SMIM21 MYO5B
OR1E2 ASPA MTATP6P3 BHLHA9 HIC1 SMG6 SERPINB7 MRPL37P1 ZNF516 CCDC11
DHX33 DERL2 CHD3 TUSC5 ACAP1 HMSD SER- RN7SL112P LINC00908 ZNF271
MIS12 NLRP1 CAMKK1 YWHAE KCTD11 PINB8 DOK6 RNU6-219P ZNF236 KLHL14

MAPRE2 RNU6-167P MIR4741 MC2R ZNF519 HMGB3P17 RPS3AP35 DEFB136 KBTBD11
ZNF397 ZSCAN30 ZNF24 RNU6-1002P RN7SL745P CCDC58P3 CEP120 OR7E8P RNA5SP258 RNA5SP256
LINC00669 RN7SKP44 RPS10P27 RNU6-170P CSNK1G3 OR7E15P DPYSL2 MTUS1 FGL1
MIR5583-1 LRRC37A7P RPL23AP77 RNU7-129P WDR36 OR7E10P ADRA1A PCM1 TNKS
RNU6-1242P RNU6-1050P RNU5A-6P LDLRAD4-AS1 RICTOR MIR5692A1 RNA5SP260 MSRA PRSS55
RPL17P45 PGDP1 DSG3 RNA5SP452 MIR5190 OSMR MIR3926-1 RPL17P33 RP1L1 C8orf74
RNU7-145P DSG2 TTR RNU6-435P MIR4526 SLC25A46 DEFB134 LINC00589 C8orf12
RN7SL233P B4GALT6 LPIN2 MYOM1 RN7SL362P SLC12A2 ZNF705D FAM183CP FAM167A
TXNDC2 SLC25A52 MYL12A CIDEA TUBB6 PGGT1B USP17L2 MIR3148 BLK GATA4
VAPA APCDD1 TRAPPC8 MYL12B AFG3L2 CCDC112 FAM86B1 MAP2K1P1 MIR597
PIGPP4 RNF125 RNU7-25P SLMO1 WDR36 DEFB130 RNU6-1218P LINC00599
RNA5SP450 RNF138 GREB1L SPIRE1 RICTOR FAM86B2 HMBOX1 C8orf49 NEIL2
ARIH2P1 RN7SKP72 ESCO1 PSMG2 SRFBP1 LONRF1 DUSP4 WRN RNU6-729P
RNU6-408P THOC1 SNRPD1 CEP76 SEH1L CSNK1G3 DEFB135 RPL5P22 PRSS51
RNU6-857P EPB41L3 ABHD3 CEP192 OSMR SDAD1P1 SCARA5 MIR4286
DSC2 DSC1 AQP4 ST8SIA5 ANKRD62 RNU6-324P MIR320C1 POTEC SLC25A46 BNIP3L NUGGC PNOC RNA5SP252
NRBF2P1 PTPN2 RNA5SP453 RNU6-1038P ANKRD30B ADAMTS19 PNMA2 NKX3-1 NKX2-6 COX6B1P4 FDFT1
MIR187 HNRNPA1P7 RNU6ATAC20P RNA5SP451 NF1P5] DMXL1 SOX7 PINX1 RN7SL293P
MIR3929 WBP11P1 GATA6-AS1 ANKRD20A5P SLC12A2 RNU6-1276P XKR6 MIR598 RNU6-1084P
MIR4318 GAREM RNU6-702P RNU6-316P PGGT1B CCDC25 LINC00529 LINC00208
RN7SKP182 MEP1B TGIF1 RHOT1P1 CCDC112 ESCO2 PBK RPL19P13 SUB1P1
RNU6-706P PTPRM YES1 DLGAP1 CYP4F35P DEFB114 SLC35G5 ENTPD4 OR7E158P
RPL7AP66 ADCYAP1 RPL31P59 TERF1P2 DEFB113 TDH MIR4288 SUMO2P16 OR7E161P
RNA5SP455 RNU1-109P IGLJCOR18 FEM1AP2 DST WDR36 RNA5SP259 RNA5SP251 CTSB NRG1-IT1 NRG1-IT2
KRT8P5 COX6CP3 RPL21P127 LONRF2P1 RICTOR FBXO16 RN7SL872P
MIR4319 GNAL RN7SL39P CXADRP3 SRFBP1 TUSC3 PAICSP4 RNA5SP262
RN7SKP26 NPIPB1P DLGAP1-AS1 GRAMD4P7 CSNK1G3 VPS37A GSR RN7SL318P RNA5SP263
RNU6-1278P CHMP1B DLGAP1-AS2 RNU6-1021P OSMR DOCK5 ELP3 RPL23AP54 NRG1-IT3
ZNF396 MPPE1 ROCK1 RNU6-120P EXOGP1 OR4K7P SLC25A46 ATP6V1B2 RN7SKP159 MIR4287
INO80C IMPA2 RAB12 OR4K8P ADAMTS19 PPM1AP1 MIR4659A RNU6-178P
C18orf21 RFWD2P1 L3MBTL4 SNX19P3 DMXL1 MSR1 GNRH1 XKR5 DEFA8P EXTL3-AS1
RPRD1A RN7SL50P C18orf64 RNU6-1210P SLC12A2 CDCA2 EBF2 DEFA9P HMBOX1-IT1
SLC39A6 THEMIS3P ARHGAP28 MIR3156-2 PGGT1B RN7SL781P DEFA10P RN7SL651P
ELP2 MOCOS AKR1B1P6 FGF7P1 CCDC112 HMGB1P23 DEFT1P MIR548H4 GU-LOP STMN4
FHOD3 TPGS2 USP14 LAMA1 RN7SL662P KCTD9 KIF13B PURG DEFT1P2
KIAA1328 COLEC12 SOGA2 LRRC30 BNIP3P3 ADAM28 LINC00681 DEFA11P TRIM35
CELF4 CETN1 CLUL1 RNU5F-3P RNU6-721P ADAMDEC1 RNU6-842P DEFA7P PTK2B
SETBP1 C18orf56 RN7SL723P GCA FSIP2 ADAM7 FZD3 KIAA1456 RPS3AP30 CHRNA2
SLC14A2 TYMS RNU6-349P RAD21L1 NRG1 RNU5A-3P GTF2E2 DLC1 MICU3 RPS3AP33 EPHX2 CLU
SLC14A1 ENOSF1 RPL6P27 SI SLC9C1 ZDHHC2 OR7E125P ZNF395
SIGLEC15 RN7SL247P RN7SL282P IQCB1 PLOD2 SMIM18 CNOT7 FAM90A15P EXTL3 INTS9
EPG5 PSTPIP2 MIR320C2 RNU6-916P STXBP5L PPP2R2A MTMR7 FAM90A3P FAM90A9P
ATP5A1 RAC1P1 RN7SL537P CENPE TECRL IL15 MTMR9 MIR3622B FAM90A4P FAM90A10P
HAUS1 WBP2P1 SLC25A51P2 ARAP2 OR7E160P RNU6-397P FAM90A20P FAM90A21P
C18orf25 RN7SL97P RPS4XP19 PRDX4P1 RNA5SP253 MIR383 FAM66B FAM90A22P
RNF165 GATA6 C18orf42 PRKRIRP9 DEFB108P3 C8orf48 SGCZ DEFB109P1B FAM90A23P
LOXHD1 CTAGE1 ZBTB14 RNU6-931P FAM66D RPS15AP24 USP17L1P OR7E157P
CIAPIN1P RBBP8 TMEM200C RN7SKP199 USP17L7 HSPA8P11 USP17L4 OR7E154P
KCTD1 CABLES1 DLGAP1-AS3 GUF1 GNPDA2 GABRG1 DEFB109P3 TUBBP1 DEFB108P2 FAM90A14P
METTL4 TMEM241 RNU6-831P ALG1L11P FAM90A2P RBPMS-AS1 HSPD1P3 FAM90A18P
NDC80 RIOK3 DLGAP1-AS4 AP1AR LCORL UBXN8 MYOM2 FAM90A16P
SMCHD1 C18orf8 NPC1 GAPDHP66 FAM85A TMEM66 CSMD1 FAM90A8P
EMILIN2 ANKRD29 DLGAP1-AS5 SNX18P23 ENPP7P12 LEPROTL1 MCPH1 ANGPT2 AGPAT5 FAM90A17P
KATNBL1P3 LAMA3 PPIAP14 GABRA2 RNA5SP254 MBOAT4 FAM90A19P
RNU6-340P TTC39C LINC00526 ADAMTS19 DEFB108P4 DCTN6 DEFB1 DEFA6 DEFB104B
CBX3P2 CABYR LINC00667 RNU6-701P ZNF705C RBPMS DEFA4 DEFA1 DEFB106B
RPS24P18 DHFRP1 NPM1P2 OSBPL1A MIR5706 FAM66A PPP2CB DEFA1B DEFB105B
FAM60BP IMPACT MIR3976 DMXL1 HNRNPKP1 MEGF10 DEFB109P1 MRPL49P2 DEFA3 DEFA5 DEFB107B
RNU6-1289P HRH4 ZNF521 IL9RP4 LINC00965 RN7SL474P ZNF705G PRR23D1
SS18 PSMA8 TAF4B RPS4XP18 ROCK1P1 PRRC1 CTXN3 FAM90A25P FGF20 RNU6-1086P SCARA3 DEFB4B PRR23D2
RNU6-1032P LDLRAD4 SKP2 NADK2 DEFB103B DEFB107A
UBE2CP2 FAM210A SRFBP1 LOX ALG1L12P NEFL NEFM SPAG11B DEFB105A
RNMT MC5R KRT8P33 ENPP7P6 MTND4P7 XPO7 DEFB106A
RPS3AP34

DEFB104A	ALG1L13P	USP17L3	FGF17 DMTN	CHMP7	WHSC1L1	ADAM18	C8orf88
SPAG11A	RN7SL178P	FAM90A11P	FAM160B2	R3HCC1	LETM2 FGFR1	ADAM2	HSPA8P13
MIR548V	RNU6-682P	WBP1LP3	NUDT18	LOXL2	RN7SKP201	IDO1 IDO2	ASNSP1
NATP	MIR4660	RPL23AP53	HR	SLC25A37	MTND6P19	C8orf4 ZMAT4	ASNSP4
RPL35P6	RNU7-55P	FAM87A	REEP4 LGI3	RNA5SP261	RNU6-533P	KCNU1	TRIM60P15
SLC7A2 PDG-	RNU6-1151P	OR4F21	SFTPC BMP1	RNU6-663P	RPL23P10	ZNF703 ER-	LINC00293
FRL ASAH1	RNU6-526P	ZNF596	PHYHIP	MTND1P6	RNA5SP264	LIN2 PROSC	MTND1P7
NAT1 NAT2	CLDN23	DEFB103A	POLR3D	MTND2P32	TPT1P8	GPR124 BRF2	RNU6-656P
PSD3 SLC18A1	MFHAS1 ERI1	DEFB4A	PIWIL2	RANP9	SMARCE1P4	RAB11FIP1	MTND6P20
SH2D4A	ADAM24P	ZNF705B	SLC39A14	RNU6-528P	RNU6-607P	GOT1L1	WDR36
CSGALNACT1	ERICH1-AS1	DOK2	PPP3CC	SNORD13	RN7SL709P	ADRB3	RICTOR
INTS10 LPL	MIR596	RNU6-892P	SORBS3	RN7SL621P	RPS20P22	EIF4EBP1	SRFBP1
RPL30P9	KBTBD11-OT1	LZTS1-AS1	PDLIM2	RN7SL457P	RNF5P1	ASH2L STAR	CSNK1G3
PPP1R3B	FBXO25	TMEM97P2	C8orf58	VENTXP5	RPL3P10	LSM1 BAG4	OSMR
FAM90A24P	TDRP ERICH1	OR6R2P	CCAR2 BIN3	FUT10 TTI2	ADAM5	DDHD2	SLC25A46
FAM90A12P	DLGAP2 CLN8	MIR320A	EGR3 PEBP4	MAK16	ADAM3A	POTEA	ADAMTS19
OR7E96P	ARHGEF10	RNU6-336P	RHOBTB2	RNF122	C8orf86 TACC1	RNU6-104P	DMXL1
MIR548I3	HSPD1P2	RN7SL303P	TNFRSF10B	DUSP26	PLEKHA2	RN7SKP41	SLC12A2
ENPP7P1	DEFB108P1	RNU4-71P	TNFRSF10C	LSM12P1	HTRA4 TM2D2	SNX18P27	PGGT1B
FAM85B	FAM66E	LZTS1	TNFRSF10D	UNC5D	ADAM9	CYP4F44P	CCDC112
FAM86B3P	USP17L8	GFRA2 NPM2	TNFRSF10A	PPAPDC1B	ADAM32	NECAB1	RFX3

D.2 Genes present in Won segments

MTND2P28	FLT1 IFIT1P1	C1QTNF9-AS1	AMER2-AS1	RPL21P109	RPLP1P13	LINC00621	CLYBL-AS1
MTATP8P1	ATP8A2P2	NUS1P3	LINC01053	RPL35AP31	SACS-AS1	BASP1P1	LINC00554
MTATP6P1	BRCA2	LINC00566	LINC01076	RPS10P21	LINC00327	NUS1P2	NDUFA12P1
MTND1P23	N4BP2L1	CYCSP33	RPL23AP69	RNA5SP32	LINC00352	PSMC1P13	ASNSP3
ZNF268	N4BP2L2	TPTE2P6	MTMR6	RPL18AP17	MTCO3P2	LINC00402	RPL15P18
RB1 UGGT2	RNU6-79P	MIPEP	FRY ZAR1L	RPL21P110	PSPC1P2	RNY1P5	PCCA-AS1
SLITRK6	PSMD10P3	C1QTNF9B	MTND5P2	RNU6-80P	RNY1P7	RPL21P108	RPS26P47
HMGN1P24	FABP5P1	SPATA13	MTND6P18	DACH1	RPL26P34	LINC00381	COX5BP6
DNAJC3	RNU6-66P	C1QTNF9	MTCYBP3	RPS21P8	IRX1P1	LINC00347	NALCN-AS1
AKAP11	RNU4-10P	PARP4	SCEL-AS1	RPS20P32	ANKRD20A10P	RIOK3P1	ARF4P3
RNF219-AS1	RNY1P8	ATXN8OS	RNY3P7	USP12-AS1	TNFRSF19	RNU6-38P	LINC00411
RN7SL810P	MARK2P12	RNU6-54P	SPTLC1P5	USP12-AS2	ATP12A	SSR1P2	SLC15A1
LINC01069	LINC00393	SOGA2P1	MIR3665	LINC00412	METTL21C	CTAGE11P	DOCK9 GPR18
LINC00446	LINC00392	KLHL1	EDNRB-AS1	SNORD102	SERPINE3	LINC01078	GPR183
SRGNP1	MZT1 BORA	RN7SKP9	RPL21P111	SNORA27	INTS6	TBC1D4-AS1	PCCA GGACT
RNY3P3	DIS3 PIBF1	H2AFZP3	LINC00331	RNU6-70P	TCEB2P1	FAM204CP	TMTC4
RPL31P54	KLF5 KLF12	UBAC2	HSPD1P8	LINC01079	RN7SL289P	TBC1D4	NALCN
TCEB1P23 ED-	PDS5B	MIR623	CCT5P2	RNY1P1	RNU6-78P	COMMD6	ITGBL1
NRB POU4F1	EIF4A1P7	HMGB3P4	NIPA2P5	RNU6-63P	RN7SL741P	UCHL3 LMO7	LINC00445
MLNR	RNU6-53P	CCR12P	BCAS2P3	USP12 RPL21	RNY1P3	RN7SKP1	RNU6-71P
PABPC3	POMP	LINC00449	MYCBP2	RASL11A	LINC00415	RPS12P24	HIST1H2APS6
AMER2	SLC46A3	TM9SF2	SCEL SLAIN1	GTF3A	ATP8A2P3	LINC01048	GAPDHP34
LINC00283	SGCG	CLYBL ZIC5	RNF219	MTIF3 LNX2	PRUNEP1	EXOSC8	TCEAL4P1
KATNBL1P1	NPM1P4	ZIC2 DNAJC3-	WASF3 GPR12	FGFR1OP2P1	PRKRIRP6	SUPT20H	ARL2BPP3
CHCHD2P8	POLR1D	AS1 DZIP1	LDHBP1	TPTE2P1	NUPL1	POSTN ALG5	NDE1P2
PAN3-AS1	RNY1P4	LINC00552	GRTP1-AS1	SLC25A15P3	ATP8A2	DOCK9-AS1	SMAD9-AS1
EEF1A1P3	PCDH8	GAS6-AS1	DCUN1D2-AS	RPL34P27	SHISA2	RPL7L1P12	LAMTOR3P1
RNU6-82P	CSNK1A1L	GAS6-AS2	RNU1-16P	RNF17 CENPJ	RNF6 HSPH1	RNU6-83P	RPL29P28
FLT3 PAN3	IFT88	LINC00454	RASA3-IT1	RPS3AP44	SDAD1P4	DOCK9-AS2	EIF4A1P5
SACS RNU6-	SLC35E1P1	LINC00452	CUL4A	WASF3-AS1	LIG4 TPP2	RPS6P23	NBEA
73P PDX1-AS1	RNU2-7P	LINC00565	LAMP1 GRTP1	CYP51A1P2	HMGA1P6	GAPDHP22	MAB21L1
LINC00543	C1QTNF9B-	TMCO3 TFDP1	ADPRHL1	POM121L13P	RNY3P4	UBAC2-AS1	DCLK1
GSX1 PDX1	AS1	ATP4B GRK1	DCUN1D2	MTUS2-AS2	LINC00362	RNY3P6	SOHLH2
ATP5EP2	ANKRD20A19P	TMEM255B	CDC16	GAPDHP69	RNU6-58P	LINC01039	CCDC169-
CDX2 URAD	IPO7P2	GAS6 RASA3	LINC00348	MTUS2-AS1	TATDN2P3	CFL1P8	SOHLH2
CCDC168	MIR2276	RBM26 UBL3	RABEPKP1	CDK8 MTUS2	IPMKP1	MIR4306	CCDC169
RN7SL272P	SPATA13-AS1	ABHD13	H3F3BP1	SLC7A1	RFESDP1	CLYBL-AS2	SPG20

SPG20OS	STK24	LINC00368	FAM58DP	METTL21EP	SOX21	LINC00366	PSME2P2
CCNA1	ANKRD26P4	ARHGEF7-AS2	SLC25A15P2	ATP6V1G1P7	RNY3P10	PRDX3P3	RNU6-60P
SERTM1	EEF1DP3	ARHGEF7-IT1	CYCSP32	RPL7P45	MOB1AP1	RNU6-56P	RAD17P2
RFXAP	LINC01073	ARHGEF7-AS1	MRPS31P2	DAOA-AS1	DDX6P2	FREM2-AS1	RNY3P2
SMAD9	FRY-AS1	LINC00354	ESRRAP1	LINC00343	TXNL1P1	FREM2	COX7CP1
MIR759	B3GALTL	LINC00404	CASC4P1	RNA5SP38	LINC00430	FHP1 DGKH	OGFOD1P1
PPIAP26	RXFP2	LINC00403	LINC00350	LINC00460	UBBP5	LINC00390	RCBTB2
LECT1	KCTD12	LINC01070	RN7SL166P	RPL35P9	LIN28AP2	SMIM2-AS1	CYSLTR2
LINC00284	OLFM4	LINC01043	ST6GALNAC4P1	LINC00551	MIR4500HG	SMIM2-IT1	FNDC3A
DGKZP1	SCAND3P1	LINC01044	CNOT4P1	LINC00443	MIR4500	LINC01071	CDADC1
STARD13-IT1	PHBP13	MIR548AR	HNRNPA1P30	ATP5G1P5	LINC00397	SMARCE1P5	CAB39L
KL STARD13	TIMM8BP1	MIR4502	PPIAP27	PPIAP24	TET1P1	LINC00407	RLIMP1
SKA3 MRP63	LINC00297	CLCP2 AR-	RPSAP54	MIR1267	RPL29P29	LINC00330	MIR320D1
ZDHHC20	LINC00572	HGEF7 TEX29	RNU6-51P	FAM155A-IT1	LINC00433	RN7SL49P	MIR621
LINC01046	LINC00544	SOX1 SPACA7	LATS2-AS1	KDELC1	LINC00560	RN7SKP3	TPTE2P5
ESRRAP2	LINC00365	TUBGCP3	RNU4-9P	BIVM BIVM-	GRPEL2P1	RNU6-69P	CYCSP34
MIPEPP3	LINC00384	C13orf35	RPS12P23	ERCC5 ERCC5	LINC01047	SNORA31	SUGT1P3
GRK6P1	LINC00385	UPF3A	IPPKP1	SLC10A2	TRIM60P13	TPT1-AS1	RGS17P1
GAPDHP52	RNU6-64P	CHAMP1	RN7SL80P	DAOA EFNB2	LINC00440	RCN1P2	MIR3168
RNA5SP25	PRDX2P1	VDAC1P12	HIST1H2BPS3	ARGLU1	SP3P	SLC25A30-AS1	RN7SL597P
ZDHHC20-IT1	LINC00427	LINC00457	FNTAP2	LINC01068	LINC01040	PPIAP25	CALM2P3
SPERT SIAH3	LINC00426	GAMTP2	RNU6-59P	LINC01038	LINC00353	RNA5SP27	MORF4L1P4
RNY3P8	LINC01058	PSMA6P4	RPS7P10	LINC00382	RPL7L1P1	TIMM9P3	RAC1P3
RNY4P27	UBE2L5P	RPL7AP61	RN7SL766P	RNU6-61P	PEX12P1	COX4I1P2	RNU6-57P
MEMO1P5	RBM22P2	RAP2A TEX30	LINC00424	HNRNPA1P31	LINC00559	LINC01055	TUBBP2
CLDN10-AS1	MFAP1P1	HSP90AB6P	NME1P1	PWWP2AP1	RNA5SP34	AKR1B1P4	OR7E36P
ABCC4	PTPN2P2	LINC00359	MTND3P1	ARF4P4	FAR1P1	SMIM2 SERP2	OR7E155P
CLDN10	LINC01066	TULP3P1	FTH1P7	DPPA3P3	KRT18P27	TSC22D1	OR7E37P
ATP11A-AS1	WDR95P KAT-	RN7SKP7	DDX39AP1	LINC00564	MIR622	NUFIP1	MIR5006
MCF2L-AS1	NAL1 HMGB1	LINC00456	RPL7AP73	RNU6-77P	LINC01049	GPALPP1	FOXO1
F10-AS1	MYCBP2-AS1	RNA5SP37	TUBA3C	HIGD1AP2 PT-	RNU6-75P	GTF2F2	MRPS31
KARSP2	MYCBP2-	HS6ST3	FEM1AP4	MAP5 GYG1P2	LINC00410	KCTD4 TPT1	SLC25A15
ATP11A	AS2 FBXL3	OXGR1	LINC00387	RNU6-67P	BRK1P2	SLC25A30	ELF1 WBP4
MCF2L F7 F10	LINC00423	MBNL2	TERF1P5	VENTXP2	LINC00380	COG3	KBTBD6
PROZ PCID2	TOMM22P3	ZMYM5 GJA3	FAM207BP	UBE2D3P4	LINC00379	FAM194B	KBTBD7
FAM155A	STARD13-AS	GJB2 GJB6	LINC00418	MTND4P1	PPIAP23	ZC3H13	MTRF1 NAA16
ZMYM2	RNU5A-	CRYL1 IL17D	GXYLT1P1	MTND5P3	MIR17HG	DIAPH3-AS1	RGCC VWA8
KRR1P1	4P RFC3	N6AMT2 XPO4	CNN2P12	LINC00333	SLITRK5	RN7SL375P	PLA2G12AP2
LINC00398	LINC00561	LINC01072	ZNF965P	LINC00375	INTS6-AS1	DIAPH3-AS2	ANKRD26P2
TEX26-AS1	LINC01034	PPIAP28	CYP4F34P	LINC00351	RPS4XP16	RNY4P28	NXT1P1
LINC00545	RN7SL571P	LINC00556	SNX18P26	RNU6-72P	RN7SL320P	LINC00434	STOML3
USPL1	BTF3P11	MIR4499	LINC00328-2P	SPRY2	SNRPGP11	TARDBPP2	PROSER1
ALOX5AP	RPL7P44	RANP8	ANKRD20A9P	SLITRK1	MIR4703	DIAPH3	NHLRC3
MEDAG	DHX9P1	MYO16-AS1	RHOT1P3	LINC00363	RNU6-65P	TDRD3	LHFP
TEX26 COG6	C13orf45	LINC00370	RNU6-55P	HNRNPA1P29	RNY1P6	COX17P1	DLEU7-AS1
COL4A2-AS2	IRG1 CLN5	LINC01067	RNU6-76P	GPC6-AS2	RN7SL413P	FKBP1AP3	RNA5SP28
COL4A2-AS1	RNA5SP33	LINC00399	SNX19P2	RNA5SP35	ATP5F1P1	LRCH1	RNASEH2B-
RPL21P107	RBM26-AS1	LINC00676	LINC00408	GPC6-AS1	CTAGE3P	SNRPGP14	AS1 GUCY1B2
LINC00567	NDFIP2-AS1	RN7SKP10	PHF2P2	RNA5SP36	FABP5P2	RNY4P30	RNA5SP29
COL4A2	NDFIP2	LINC00396	LINC00442	RN7SL585P	MRPS31P5	CTAGE10P	LINC00371
RAB20 CARKD	RNU1-24P	RN7SL783P	USP24P1	LINC00391	TPTE2P2	RNY4P9	RPL5P31
CARS2 ING1	HMGB3P7	MYO16 IRS2	GTF2IP3	SOX21-AS1	RNY4P24	MIR3613	SLMO2P2
LINC00346	MIR2681	COL4A1	RNA5SP24	BRD7P5	LINC00345	DLEU2 MIR16-	MIR5693
ANKRD10	LIPT1P1	RN7SL164P	CENPIP1	RPL21P112	TPTE2P3	1 RPL18P10	DLEU7
FTLP8	RNY1P2	AMMECR1LP1	RNU6-52P	LINC00557	MRPS31P4	SETDB2	RNASEH2B
RN7SKP8	MIR4705	RNA5SP26	SMPD4P2	RNU6-62P	WDFY2	PHF11	FAM124A
MIR3170	RPL39P29	HSPD1P9	PSPC1P1	RNU4ATAC3P	DHRS12	RCBTB1	CHCHD2P11
FARP1-AS1	FGF14-IT1	LINC00571	ANKRD26P3	FABP5P4	CCDC70	ARL11 EBPL	RNA5SP31
STK24-AS1	FGF14-AS1	TRPC4	MRPL3P1	GPC5-AS2	ATP7B ALG11	KPNA3	EIF4A1P6
NUS1P4	FGF14-AS2	UFM1 TPTE2	LINC00421	GPC5-IT1	UTP14C NEK5	SPRYD7	LINC00378
CYCSP35	LINC00555	MPHOSPH8	RNA5SP39	MIR548AS	NEK3 THSD1	TRIM13	RNY3P5
CALM2P4	FGF14 ESD	PSPC1	HCFC2P1	GPC5-AS1	VPS36 CKAP2	KCNRG	RNY4P31
RN7SL60P	ANKRD10-IT1	LATS2 SAP18	MYO16-AS2	GPC5 GPC6	HNRNPA1L2	DLEU1	MIR3169
IPO5 FARP1	PARP1P1	MICU2 FGF9	TNFSF13B	DCT TGDS	SUGT1	LINC01077	RAC1P8
RNF113B	LINC00431	PARP4P2	RNY5P8	GPR180	LINC00437	LINC00462	LINC00358

LINC01075 RN7SKP2 PRR20E MIR548O2 TOMM34 SNAP23P RPS2P1 TTLL9
LINC01074 LINC00598 PCDH17 RPS3P2 STK4 KCNS1 ARPC3P1 XPOTP1 PDRG1 XKR7
LINC00459 RN7SL515P ZNF962P RN7SKP173 RNU6-407P ZFAS1 CDC42P1 DKKL1P1
RPL32P28 RPS28P8 BNIP3P7 RN7SL116P RNA5SP483 SNORD12C ITCH-AS1 RPL31P3
SQSTM1P1 MAPK6PS3 LINC00349 NPM1P19 MMP24-AS1 SNORD12B ITCH-IT1 HAUS6P2
LINC00448 FABP3P2 LONRF2P2 ATG3P1 MT1P3 SNORD12 MIR644A RNA5SP480
LINC00376 LINC01050 LINC00388 RN7SL680P FAM83C-AS1 RN7SL197P FDX1P1 DEFB117
LINC00395 LINC00428 TDRD15 FSIP2 RN7SL194P MIR1289-1 SNRPFP1 CCM2L HCK DEFB122
OR7E156P ZDHHC4P1 TTN-AS1 HSPE1P1 RPL36P4 RNU6-919P TM9SF4 LINC00028
RNU6-81P LINC00400 TTN EPPIN- RNA5SP484 FER1L4 RNU7-72P PLAGL2 HM13-IT1
PPP1R2P10 RPL36P19 WFDC6 RNU2-52P RPL37P1 KRT18P4 POFUT1 HM13-AS1
OR7E104P ENOX1-AS2 WFDC8 RPL23AP81 RN7SKP271 RNU6-147P KIF3B ASXL1 RNU6-384P
NFYAP1 ENOX1-AS1 WFDC9 RN7SL615P RNU6-759P TRERNA1 C20orf112 MIR3193
LINC00355 CPB2-AS1 WFDC10A RPL12P11 RPF2P1 RN7SL636P COMMD7 ERP29P1
LGMNP1 RN7SL288P WFDC11 RNU6-1018P RNU6-1166P COX6CP2 DNMT3B MRPS33P4
STARP1 RN7SKP5 WFDC10B MROH8 COX7BP2 RN7SL672P MAPRE1 RPL36P1
HNRNPA3P5 FAM206BP WFDC13 RPN2 GHRH HIGD1AP16 MIR645 EFCAB8 RN7SKP184
LINC01052 LINC00563 SPINT4 MANBAL SRC RNU4-40P MIR1302-5 SUN5 BPIFB2 PPIAP10
MIR548X2 RNU2-6P WFDC3 DNT- BLCAP NNAT RNU6-937P RPL36P2 BPIFB6 RNU7-14P
MIR4704 RNU6-68P TIP1 UBE2C CTNNBL1 LINC00657 TMSB4XP6 BPIFB3 SUMO1P1
TRIM60P19 PPP1R2P4 TNNC2 VSTM2L TTI1 HMGB3P2 PSMD10P1 BPIFB4 MIR4756
PCDH9-AS1 OR7E101P SNX21 ACOT8 RPRD1B HNRNPA3P2 RPSAP1 BPIFA2 TSHZ2 ZNF217
RNU7-87P HTR2A-AS1 ZSWIM3 TGM2 RN7SL156P MIR3194 BPIFA3 BCAS1
PCDH9-AS2 GNG5P5 ZSWIM1 KIAA1755 RPS3AP3 RNU6-347P BPIFA1 CYP24A1
PCDH9-AS3 RN7SL700P SPATA25 BPI LBP RAL- RPS27AP3 RNU7-6P BPIFB1 PFDN4 DOK5
PCDH9-AS4 NAP1L4P3 NEURL2 CTSA GAPB ADIG TRPC4AP RN7SL603P CDK5RAP1 DEFB115
RPSAP53 RPL27AP8 MRPL5P2 ARHGAP40 EDEM2 PREX1 SNTA1 DEFB116
LINC00364 LINC00444 HSPD1P21 SLC32A1 PROCR ARFGEF2 CBFA2T2 DEFB118
BCRP9 LINC00562 RNA5SP485 ACTR5 MMP24 EIF6 CSE1L STAU1 NECAB3 DEFB119
NPM1P22 LINC01065 CCNB1IP1P2 PPP1R16B FAM83C DDX27 ZNFX1 C20orf144 DEFB121
OR7E111P PCDH8P1 RPS2P7 FAM83D UQCC1 KCNB1 PTGIS ACTL10 DEFB123
OR7E33P RN7SL618P IR3617 DHX35 MAFB GDF5OS B4GALT5 E2F1 PXMP4 DEFB124
ELL2P3 LINC00558 HNRNPA1P3 TOP1 PLCG1 GDF5 CEP250 SLC9A8 ZNF341 REM1 HM13
HNRNPA1P18 LINC00458 SPINT5P ZHX3 LPIN3 C20orf173 ER- SPATA2 CHMP4B ID1 COX4I2
RPL37P21 RPL13AP25 RNU6ATAC38P EMILIN3 GIC3 SPAG4 RNF114 SNAI1 RALY BCL2L1
RPS3AP52 MIR5007 FTLP1 CHD6 PTPRT CPNE1 RBM12 TMEM189- EIF2S2 ASIP C20orf166-
RPL12P34 HNF4GP1 RPL13P2 SEMG2 NFS1 ROMO1 UBE2V1 AHCY ITCH AS1 MIR1-1
RN7SL761P SPATA2P1 PLTP PCIF1 RNU6-1251P RBM39 PHF20 UBE2V1 ZNF663P MIR133A2
ZDHHC20P4 RN7SKP6 ZNF335 MMP9 RPL27AP SCAND1 TMEM189 MKRN7P RPL7P3
SNRPFP3 PRR20FP SLC12A5 RNU6-639P CNBD2 CEBPB PTPN1 ZNF840 LINC00686
LINC00383 MTCO2P3 NCOA5 RN7SL443P EPB41L1 FAM65C RN7SKP33 LINC00659
LINC00401 SLC25A5P4 CD40 CDH22 OSER1-AS1 AAR2 DLGAP4 PARD6B GAPDHP54 OGFR-AS1
SRSF1P1 RPL31P53 SLC35C2 HNF4A-AS1 MYL9 TGIF2 BCAS4 ADNP MIR3616 LAMA5 RPS21
PCDH20 RNA5SP30 ELMO2 SLPI MIR3646 TGIF2- DPM1 MOCS3 RN7SL243P CABLES2
PCDH9 RNU6- LINC00374 MATN4 RBPJL RPL37AP1 C20orf24 KCNG1 RPL35AP RBBP8NL
74P KARSP1 RNY4P29 SDC4 06] RN7SL31P C20orf24 SLA2 NFATC2 RNU6-497P GATA5
SUCLA2-AS1 CTAGE16P SYS1 SYS1- STK4-AS1 NDRG3 DSN1 ATP9A SALL4 RNU6-563P C20orf166
MED4-AS1 DNAJA1P1 DBNDD2 SRSF6 SOGA1 TLDC2 ZFP64 ZNF334 RNU7-173P SLCO4A1
POLR2KP2 HMGN2P39 TP53TG5 L3MBTL1 SAMHD1 RNA5SP482 SRMP1 NTSR1 MR-
LINC00441 POLR3KP1 DBNDD2 SGK2 IFT52 RBL1 WFDC5 RSL24D1P6 RNA5SP486 GBP OGFR
PPP1R26P1 RPP40P2 PIGT WFDC2 MYBL2 WFDC12 TSPY26P RNU7-92P COL9A3
PCNPP5 RNU7-88P SPINT3 GTSF1L TOX2 PI3 SEMG1 MIR1825 LINC00494 LINC00029
ST13P4 TNFSF11 WFDC6 EPPIN JPH2 OSER1 HMGB3P1 C20orf203 RNU7-144P LINC01056
RPL34P26 FAM216B EP- RPL7AP14 GDAP1L1 MIR499A BAK1P1 OCSTAMP HAR1B
SUCLA2 STI1 DNAJC15 PPIAP3 FITM2 DYNLRB1 HDHD1P3 SLC13A3 HAR1A
NUDT15 ENOX1 LINC00489 R3HDML MAP1LC3A SOCS2P1 TP53RK MIR124-3
MED4 ITM2B CCDC122 RN7SKP185 HNF4A PIGU NCOA6 BPIFA4P SLC2A10 MIR3196
LPAR6 LACC1 CPB2 RN7SL237P C20orf62 TP53INP2 RPL12P3 EYA2 MIR4326
MIR4305 LCP1 LRRC63 SNHG17 TTPAL SER- GGT7 ACSS2 BPIFB5P ZMYND8 RNU6-994P
RNY4P14 KIAA0226L SNORA71B INC3 PKIG GSS MYH7B BPIFB9P NCOA3 SULF2 DIDO1 GID8
CDKN2AIPNLP3 HTR2A SNORA71A ADA WISP2 RN7SKP100 RN7SL511P RNU1-94P SLC17A9
AZU1P1 PRR20A SNORA71C KCNK15 RN7SL666P RPL31P2 RNA5SP481 BHLHE23
LINC00332 PRR20B SNORA71D RIMS4 PPIAP21 TPM3P2 TPX2 MYLK2 YTHDF1
RPL17P51 PRR20C SNHG11 YWHAB RNU6-743P PIGPP3 FOXS1 BIRC7
RNY3P9 PRR20D SNORA60 PABPC1L EIF4EBP2P1 MIR4755 DUSP15 NKAIN4

ARFGAP1 SPO11 RAE1 CST9LP1 SNORD17 RN7SL498P RNU6-1019P MOSPD3 RNA5SP233
COL20A1 MTRNR2L3 CST2P1 CSTP1 PTMAP3 TARDBPP1 UBE2V1P1 TFR2 ACTL6B PHBP15
CHRNA4 RBM38 POM121L3P RPL15P1 CASC20 SF3A3P1 GNB2 GIGYF1 STAG3L2
KCNQ2 MTND1P9 RNU1-23P RNU7-137P RN7SL547P MIR103A2 POP7 EPO PMS2P5
MIR4758 EDN3 GAPDHP53 SNX5 OVOL2 TMX4 PHKBP1 FTLP3 ZAN EPHB4 NCF1C
DPH3P1 PHACTR3 NAPB CSTL1 MGME1 PLCB1-IT1 RPL21P2 SLC12A9 GTF2IP1
ARF4P2 PIEZO1P1 CST11 CST8 PET117 RNU105B RPS4XP2 TRIP6 SRRT PHBP6 LIMK1
TCFL5 RNU1-134P MIR941-2 CICP4 CST9L CST9 DYNLT3P1 SNAP25-AS1 DEFB125 UFSP1 ACHE EIF4H LAT2
LINC00266-1 CST3 CST4 RPS27AP2 HIGD1AP15 DEFB126 MUC3A RFC2 CLIP2
MIR941-3 CTCFL PCK1 CST1 CST2 RN7SKP69 RPL23AP6 DEFB127 MUC12 GTF2IRD1
MIR941-1 ZBP1 PMEPA1 CST5 GGTLC1 RNU6-192P PLCB1 PLCB4 DEFB128 MUC17 GTF2I NCF1
MIR1914 C20orf85 RN7SKP140 PCSK2 BFSP1 LAMP5 PAK7 DEFB129 TRIM56 SER-PINE1 AP1S1 GTF2IRD2
MIR647 ANKRD60 GSTM3P1 DSTN RRBP1 ANKEF1 DEFB132 GATSL1
LINC00176 PPP4R1L NKX2-2-AS1 BANF2 SNAP25 C20orf96 VGF NAT16 WBSCR16
EEF1A2 RAB22A SLC25A6P1 PPIAP17 MKKS ZCCHC3 MOGAT3 GTF2IRD2B
PPDPF HMGB1P1 RPL41P1 RPLP0P1 SLX4IP JAG1 SOX12 NRSN2 PLOD3 ZN-HIT1 CLDN15 GATSL2
PTK6 SRMS MIR4532 LINC00261 RNU6-27P C20orf187 TRIB3 RBCK1 USP42 CYTH3
C20orf195 APCDD1L-AS1 NKX2-2 PAX1 RNU1-131P BTBD3 TBC1D20 FIS1 RABL5 COG5 GPR22
HELZ2 PIEZO1P2 FOXA2 INSM1 MACROD2 SPTLC3 ISM1 CSNK2A1 COL26A1 TUBG1P
GMEB2 MIR296 EIF4E2P1 KIF16B TASP1 ESF1 TCF15 SRXN1 MYL10 POLR2J4
STMN3 MIR298 GNAS-AS1 VAPB RN7SL607P MRPS11P1 SNRPB2 OTOR NDUFAF5 SEL1L2 SCRT2 SLC52A3 CUX1 SH2B2 PRKRIP1 URGCP-MRPS24
RTEL1 RTEL1-TNFRSF6B APCDD1L STX16 STX16-NPEPL1 LINC00237 RPL24P2 SDAD1P2 FAT1P1 FLRT3 IDH3B EBF4 CPXM1 FAM110A ANGPT4 ORAI2 AL-KBH4 LRWD1 URGCP UBE2D4
TNFRSF6B ARFRP1 NPEPL1 RNA5SP477 RPS11P1 TMEM239 RSPO4 PSMF1 POLR2J AKAP9
ZGPAT LIME1 NPEPL1 RPS15AP1 PGAM3P C20orf141 TMEM74B RASA4B ZNF655
SLC2A4RG GNAS CTSZ KRT18P3 LINC00687 PCED1A C20orf202 POLR2J3 SP-DYE2 RASA4 RPL23P8 LINC01162
ZBTB46 TUBB1 ATP5E LINC00656 RN7SKP111 VPS16 PTPRA RAD21L1
ABHD16B SLMO2 RNA5SP478 PA2G4P2 GNRH2 SNPH SD-CBP2 FKBP1A UPK3BL RPS26P30
TPD52L2 ZNF831 RNA5SP479 ISM1-AS1 MRPS26 OXT POLR2J2 RN7SL542P
DNAJC5 MRPS16P2 RALGAPA2 GAPDHP2 AVP UBOX5 NSFL1C SPDYE2B ASS1P11
UCKL1 NELFCD SSTR4 RNU6-278P FASTKD5 SIRPB2 SIRPD FAM185A RNU1-15P
ZNF512B SYCP2 RNU7-141P FAM217B THBD CD93 NXT1 GZF1 RPS3P1 DDRGK1 SIRPB1 SIRPG CYP3A52P SP8 CRHR2
SAMD10 AIMP1P1 ITPA SLC4A11 SIRPA PDYN AZGP1P1 INMT INMT-FAM188B
PRPF6 SOX18 PPP1R3D NCOR1P1 RN7SL864P C20orf194 STK35 TGM3 MIR25 MIR93
TCEA2 CDH26 MIR663A RNU6-228P ATRN GFRA4 TGM6 SNRPB MIR106B FAM188B
RGS19 OPRL1 C20orf197 RPL17P1 MACROD2-IT1 ADAM33 ZNF343 RPL7P60 AQP1 GHRHR
C20orf201 MIR646 RN7SL690P RPS10P2 SIGLEC1 TMC2 NOP56 MIR4658 ADCYAP1R1
NPBWR2 MTCO2P1 SLC24A3 MACROD2-AS1 RNU6-1159P HSPA12B NAMPT PMS2P1 CASD1 SGCE
MYT1 OSBPL2 MIR4533 RIN2 NAA20 C20orf27 RN7SL161P CYP3A4 RASA4CP
ADRM1 CSTF1 MIR548AG2 CRNKL1 SPEF1 CENPB RN7SL416P CYP3A43 SPDYE1
CASS4 RT-FDC1 GCNT7 MLLT10P1 C20orf26 RNU6-115P CDC25B PCOLCE-AS1 STAG3L5P LINC00957
FRG1B GINS1 RN7SL14P RNA5SP475 AP5S1 MAVS RN7SL750P STAG3L5P-PVRIG2P-PILRB MIR4649
FAM209A XRN2 RNU6-1257P PPIAP2 RNU2-56P ENSAP1 PANK2 RN7SL549P RNA5SP230
FAM209B RN7SKP74 RASSF2 RNF24 SMOX RN7SKP54 RNU6-1097P
TFAP2C BMP7 RNU6ATAC17P RNY4P11 SLC23A2 ADRA1D MIR4653 PVRIG2P DBNL PGAM2
RNU4ATAC7P ZNF337-AS1 LINC00851 TMEM230 PRNP DGAT2L7P OR2AE1 POLM AEBP1
RPL12P4 RN7SL594P GCNT1P1 PCNA CDS2 PRND PRNT RNU6-1104P TRIM4 GJC3 POLD2 MYL7
RNA5SP487 MED28P7 RNA5SP476 PROKR2 ACTG1P3 AZGP1P2 AZGP1 GCK YKT6
RPL39P VN1R108P MIR3192 GPCPD1 SDCBP2-AS1 LINC01007 ZKSCAN1 CAMK2B
RPS4XP3 SYNDIG1 RPL21P3 C20orf196 RNU6-917P MIR4285 ZSCAN21 NUDCD3
RNU6-1146P CST7 APMAP RPS19P1 CHGB TRMT6 SIRPB3P SPDYE6 ZNF3 COPS6 NPC1L1
RN7SL170P ACSS1 LINC00493 MCM8 CRLS1 RN7SL561P MIR548O MCM7 AP4M1 DDX56
PTMAP6 VSX1 EN-TPD6 PYGB RN7SL638P LRRN4 RPL7P2 MIR5090 TAF6 CNPY4 TMED4 OGDH
RNU6-929P RNU6ATAC34P FERMT1 SNORD119 MIR4467 MBLAC1 KCP FLNC
MIR4325 ABHD12 NINL DUXAP7 BMP2 MIR1292 RASA4DP LAMTOR4 ATP6V1F
CBLN4 MC3R NANP ZNF337 EEF1A1P34 RN7SL514P SNORD110 ZCWPW1 C7orf43 IRF5 MIR4657
FAM210B FAM182B LINC00652 RNA5SP474 SNORA51 MEPCE GAL3ST4 SNHG15
AURKA CDH4 CFTRP1 CSRP2BP UBE2D3P1 SNORD86 PPP1R35 GPC2 STAG3 SNORA5A
TAF4 LSM14B PLK1S1 ZNF133 DZ-ANK1 POLR3F RNA5-8SP7 SNORD56 C7orf61 GATS PVRIG SNORA5C
PSMA7 ZNF877P LINC00658 SNORD57 TSC22D4 SPDYE3 SNORA5B
SS18L1 MTG2 NKX2-4 RBBP9 LINC00654 RPL19P1 NYAP1 AGFG2 PILRB PILRA ELK1P1
MIR1257 FAM182A SEC23B DTD1 HAO1 RN7SL555P LRCH4 SAP25 COL28A1 SEPT7P2
HRH3 CST12P C20orf78 RPS18P1 UBOX5-AS1 FBXO24 MIOS MIR590 RNU6-241P
PCMTD2 CST13P SCP2D1 RNU1-55P RN7SL839P PCOLCE RNU6-1070P CICP20 RNU6-

326P ZMIZ2
PPIA H2AFV
PURB MYO1G
CCM2 NACAD
TBRG4
RAMP3
ADCY1
IGFBP1
IGFBP3
SP4 PHF14
THSD7A
NPM1P11
RBAK-
RBAKDN
RBAK TWIST1
FERD3L
CICP22 WIPI2
SLC29A4
TNRC18
FBXL18
ACTB FSCN1
RNF216
OCM CCZ1
RSPH10B
PMS2 AIMP2
EIF2AK1
ANKRD61
FAM220A
RAC1 DAGLB
KDELR2
GRID2IP
ZDHHC4
C7orf26
ZNF853
RBAKDN
OR10AH1P
ZNF890P
RNU6-215P
MIR589
RNF216-IT1
ZNF815P
RN7SL556P
RN7SL851P
RNU6-218P
ZNF316
CCDC136
ABCB5
ABCA13
TRIL TNPO3
TSEN15P3
MIR148A
MPP6 DFNA5
OSBPL3
CYCS C7orf31
NPVF NFE2L3

HNRNPA2B1
CBX3 SNX10
RNA5SP228
RNU6-1103P
SUMO2P14
RPL7AP38
HMGB3P20
TPM3P4
NHP2P2
HOTAIRM1
HOXA-AS2
HOXA-AS3
HOXA10-AS
MIR196B
HOXA11-AS
HOTTIP EVX1-
AS RPL35P4
HNRNPA1P73
EIF4HP1
PSMC1P2 NPY
KIAA0087
C7orf71
SKAP2 HOXA1
HOXA2
HOXA3
HOXA4
HOXA5
HOXA6
HOXA7
HOXA9
HOXA10
HOXA11
HOXA13 EVX1
HIBADH
TAX1BP1
CARD11 SDK1
FOXK1 AP5Z1
RADIL PA-
POLB MMD2
RN7SKP130
CYP3A54P
MIR4656
RNF216P1
RN7SL7P
CDK6 SAMD9
SAMD9L
HEPACAM2
RN7SL365P
RNU6-979P
JAZF1-AS1
JAZF1 CCL26
CCL24
RHBDD2 POR
TMEM120A
STYXL1

MDH2 SRRM3
HSPB1
YWHAG
SRCRB4D ZP3
DTX2 UPK3B
MIR4284
RN7SL265P
LINC00035
PMS2P10
PMS2P2
STAG3L1
NSUN5P1
PMS2P3
RN7SL642P
MIR4651
RPL7L1P3
SNORA14A
RNU6-863P
RN7SL212P
FDPSP2 TBL2
MLXIPL
VPS37D
DNAJC30
WBSCR22
STX1A
ABHD11
CLDN3
CLDN4
WBSCR27
WBSCR28
ELN TRIM73
POM121C
HIP1 CALU
OPN1SW
POMZP3
FDPSP7
DTX2P1
DTX2P1-
UPK3BP1-
PMS2P11
UPK3BP1
PMS2P11
SPDYE9P
PMS2P9
BCL7B
FAM185BP
BAZ1B SNX13
AHR UNCX
MICALL2
INTS1 MAFK
TMEM184A
PSMG3
CYP2W1
C7orf50
GPR146

GPER1
ZFAND2A
FAM20C
PDGFA
PRKAR1B
HEATR2
SUN1 GET4
ADAP1 COX19
MIR339
ELFN1
MAD1L1
PSMG3-AS1
TFAMP1
MIR4655
FTSJ2 NUDT1
SNX8 EIF3B
CHST12
LFNG BRAT1
IQCE TTYH3
AMZ1 GNA12
GRIFIN
MIR4648
PRPS1L1
HDAC9
RNMTL1P2
CCDC132
AGR3 ZNF12
RSPH10B2
CCZ1B
PMS2CL
MTBP ATAD2
SNX16
PHF20L1
MRPL13
SNTB1 DUTP2
ADCY8
MSC ZHX2
DERL1 EFR3A
ARFGEF1
TMEM71
HNRNPA1P36
RNU6-628P
RNU6-875P
MIR548AA1
ZHX1-
C8ORF76
ZHX1 FAM83A-
AS1 MIR4663
UBA52P5
TBC1D31
FAM83A
C8orf76
RN7SL590P
RNU6-442P
LINC00861

SOD1P3
KNOP1P5
RFPL4AP5
RNU6-869P
RNU11-4P
PCAT1 PCAT2
PRNCR1
CASC19
CCAT1 CASC8
CASC11 NS-
MCE2 TRIB1
FAM84B
POU5F1B
MYC NDUFB9
MTSS1 STAU2
TAF2 SLC45A4
GPR20
PTP4A3
MROH5
TSNARE1
HNRNPA1P38
MIR4472-1
LINC00051
MROH4P JRK
ZNHIT1P1
CDC42P3
AK3P2 BAI1
RNU6-220P
RHPN1-AS1
ARC PSCA
LY6K THEM6
SLURP1
LYPD2 LYNX1
LY6D GML
CYP11B1
CYP11B2
LY6E C8orf31
LY6H GPI-
HBP1 ZFP41
GLI4 ZNF696
TOP1MT
RHPN1
CYHR1
KIFC2 FOXH1
PPP1R16A
GPT MFSD3
RECQL4
LRRC14
LRRC24
C8orf82
ARHGAP39
RNU7-109P
MIR4664
FAM83H-AS1
MIR937

MIR661 OP-
LAH TSSK5P1
MIR939
MIR1234
MAFA ZC3H3
GSDMD
MROH6
NAPRT1
EEF1D TIGD5
PYCRL TSTA3
ZNF623
ZNF707
CCDC166
MAPK15
FAM83H
SCRIB PUF60
NRBP2 EPPK1
PLEC PARP10
GRINA
SPATC1
EXOSC4
GPAA1 CYC1
SHARPIN
MAF1
KIAA1875
FAM203A
FAM203B
BOP1 SCXA
HSF1 DGAT1
SCRT1
TMEM249
SLC52A2
FBXL6
ADCK5 CPSF1
SLC39A4
VPS28 TONSL
DENND3
RAD54B
FSBP ZNF251
ZNF34 RPL8
ZNF517 ZNF7
COMMD5
ZNF250 ZNF16
RN7SL395P
MIR4662A
LINC00964
KIAA0196-AS1
RN7SL329P
ZNF572 SQLE
KIAA0196
LRRC6 PTK2
RNA5SP278
ZNF252P
TMED10P1

ZNF252P-AS1
C8orf33
RPL35AP19
HAS2-AS1
MRPS36P3
LINC01151
RNY4P5
HMGB1P19
HAS2 IM-
PDH1P6
RN7SKP155
FER1L6-AS1
FER1L6-AS2
ARF1P3
RNU6-756P
RNF139-AS1
PVT1 MIR1205
RNU1-106P
RNU4-25P
MIR1207
MIR1208
RN7SKP226
LINC00977
RN7SKP206
CCDC26
MIR3686
RNU7-181P
MIR5194
RNU6-1255P
WDYHV1
FBXO32
KLHL38
ANXA13
FAM91A1
FER1L6
TMEM65
TRMT12
RNF139
TATDN1
TMEM75
GSDMC
FAM49B
ASAP1
RN7SL396P
RNA5SP277
HPYR1 DSCC1
DEPTOR
COL14A1
OC90 HHLA1
KCNQ3 RP1
EXT1 NOV
PKHD1L1
CSMD3
RNA5SP266
LINC00588

RNU6-596P
RPL30P10
RNU7-174P
IMPAD1
FAM110B
CYP7A1
SDCBP EIF3H
UTP23 RAD21
UBXN2B
AGO2 RAD21-
AS1 MIR3610
RN7SL228P
RN7SL826P
RPS10P16
RPS26P35
RNU6-12P
MAL2
RN7SKP153
WISP1-OT1
ST13P6
MTND2P7
ZFAT-AS1
MIR30B
MIR30D
MAPRE1P1
RNU1-35P
RNU6-144P
CASC7
MIR151A
AARD
SLC30A8
MED30
SAMD12
TNFRSF11B
COLEC10
ENPP2 TG
SLA WISP1
NDRG1
ST3GAL1
ZFAT KH-
DRBS3
FAM135B
COL22A1
KCNK9
TRAPPC9
C8orf17
CHRAC1
LINC00536
RNA5SP276
TRPS1