

RESEARCH ARTICLE

# A Meta-Analysis of Retinoblastoma Copy Numbers Refines the List of Possible Driver Genes Involved in Tumor Progression

Irsan E. Kooi<sup>1</sup>, Berber M. Mol<sup>1</sup>, Maarten P. G. Massink<sup>2</sup>, Marcus C. de Jong<sup>3</sup>, Pim de Graaf<sup>3</sup>, Paul van der Valk<sup>4</sup>, Hanne Meijers-Heijboer<sup>1</sup>, Gertjan J. L. Kaspers<sup>5</sup>, Annette C. Moll<sup>6</sup>, Hein te Riele<sup>7</sup>, Jacqueline Cloos<sup>5,8</sup>, Josephine C. Dorsman<sup>1\*</sup>

**1** Department of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands, **2** Department of Bio-medical Genetics, University Medical center Utrecht, Utrecht, The Netherlands, **3** Department of Radiology and Nuclear Medicine, VU University Medical Center, Amsterdam, The Netherlands, **4** Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands, **5** Department of Pediatric Oncology/Hematology, VU University Medical Center, Amsterdam, The Netherlands, **6** Department of Ophthalmology, VU University Medical Center, Amsterdam, the Netherlands, **7** Division of Biological Stress Response, Netherlands Cancer Institute, Amsterdam, The Netherlands, **8** Department of Hematology, VU University Medical Center, Amsterdam, The Netherlands

\* [jc.dorsman@vumc.nl](mailto:jc.dorsman@vumc.nl)



OPEN ACCESS

**Citation:** Kooi IE, Mol BM, Massink MPG, de Jong MC, de Graaf P, van der Valk P, et al. (2016) A Meta-Analysis of Retinoblastoma Copy Numbers Refines the List of Possible Driver Genes Involved in Tumor Progression. PLoS ONE 11(4): e0153323. doi:10.1371/journal.pone.0153323

**Editor:** Ralf Krahe, University of Texas MD Anderson Cancer Center, UNITED STATES

**Received:** October 9, 2015

**Accepted:** March 28, 2016

**Published:** April 26, 2016

**Copyright:** © 2016 Kooi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files. Also gene expression data are available at from NCBI GEO using accession numbers GSE59983 (primary samples) and GSE77094 (cell lines). DNA copy number data are available at the European Genome-phenome Archive under accession number EGAS00001001715.

**Funding:** The research was supported by grants from VUmc CCA (VU University Medical Center Cancer Center Amsterdam foundation) grant number CCA20083-05 and KIKA (Dutch Children Cancer-

## Abstract

### Background

While *RB1* loss initiates retinoblastoma development, additional somatic copy number alterations (SCNAs) can drive tumor progression. Although SCNAs have been identified with good concordance between studies at a cytoband resolution, accurate identification of single genes for all recurrent SCNAs is still challenging. This study presents a comprehensive meta-analysis of genome-wide SCNAs integrated with gene expression profiling data, narrowing down the list of plausible retinoblastoma driver genes.

### Methods

We performed SCNA profiling of 45 primary retinoblastoma samples and eight retinoblastoma cell lines by high-resolution microarrays. We combined our data with genomic, clinical and histopathological data of ten published genome-wide SCNA studies, which strongly enhanced the power of our analyses (N = 310).

### Results

Comprehensive recurrence analysis of SCNAs in all studies integrated with gene expression data allowed us to reduce candidate gene lists for 1q, 2p, 6p, 7q and 13q to a limited gene set. Besides the well-established driver genes *RB1* (13q-loss) and *MYCN* (2p-gain) we identified *CRB1* and *NEK7* (1q-gain), *SOX4* (6p-gain) and *NUP205* (7q-gain) as novel retinoblastoma driver candidates. Depending on the sample subset and algorithms used, alternative candidates were identified including *MIR181* (1q-gain) and *DEK* (6p gain). Remarkably, our study showed that copy number gains rarely exceeded change of one

Free foundation) grant number 98. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

copy, even in pure tumor samples with 100% homozygosity at the *RB1* locus (N = 34), which is indicative for intra-tumor heterogeneity. In addition, profound between-tumor variability was observed that was associated with age at diagnosis and differentiation grades.

## Interpretation

Since focal alterations at commonly altered chromosome regions were rare except for 2p24.3 (*MYCN*), further functional validation of the oncogenic potential of the described candidate genes is now required. For further investigations, our study provides a refined and revised set of candidate retinoblastoma driver genes.

## Introduction

Retinoblastoma is a pediatric cancer of the retina. Although the disease is relatively rare accounting for 2% of childhood cancers [1], retinoblastoma is the most common intra-ocular malignancy in children [2]. Retinoblastoma development is initiated by two sequential hits [3] of *RB1* (*RB1*<sup>-/-</sup> patients) and in few cases by amplification of *MYCN* (*RB1*<sup>+/+</sup>*MYCN*<sup>A</sup> patients) [4]. Hereditary patients carry a deleterious germ line mutation in one *RB1* allele and therefore only require one somatic mutation in the wild type *RB1* allele for retinoblastoma to develop while non-hereditary patients require two somatic mutations in *RB1*. However, while bi-allelic inactivation of *RB1* can cause benign retinoma lesions, additional genetic alterations can be required for progression to retinoblastoma [5].

Several studies aimed at identifying genetic alterations subsequent to *RB1* inactivation have yielded useful insights. Early comparative genome hybridization (CGH) [6–10], array-CGH [11–13] and more recent single nucleotide polymorphism (SNP)-array profiling studies [14–16] showed that retinoblastoma genomes frequently (>10%) contain somatic copy number alterations (SCNAs) including gains at chromosomal arms 1q, 2p, 6p, 13q and 19q and losses at 13q, 16q and 17p. Except for the 2p-arm (2p24.3, *MYCN*), focal SCNAs (< 3 Mb) are rare [15] complicating the refinement of minimal regions of gain/loss (MR[G/L]) to a single gene level. Moreover, sample sizes of published studies were small which further impeded identification of driver genes within recurrent SCNA-regions.

In addition, it has been demonstrated that there is profound variability in the total amount of genomic disruption by SCNAs between retinoblastoma tumors [15]. In several studies it was discussed whether and how the extent of genomic disruption relates to clinical and histopathological variables. However, due to strong connectivity between the independent variables (like age at diagnosis, heredity, laterality and differentiation) and small sample sizes, explanations for the variability in genomic disruption in retinoblastoma were inconclusive due to limited power.

Our study aims to refine the set of putative driver genes of recurrent SCNAs and get insight into the variability in genomic disruption. Data from high-resolution genome-wide SNP-arrays of 45 human retinoblastoma samples matched with peripheral blood DNA were used and complemented with clinical and histopathological features. In order to increase the power of our study, results were analyzed together with results of ten published genome-wide SCNA-profiling retinoblastoma studies [6–11,13,15–17] adding up to a considerable number of 310 primary retinoblastoma samples.

## Results

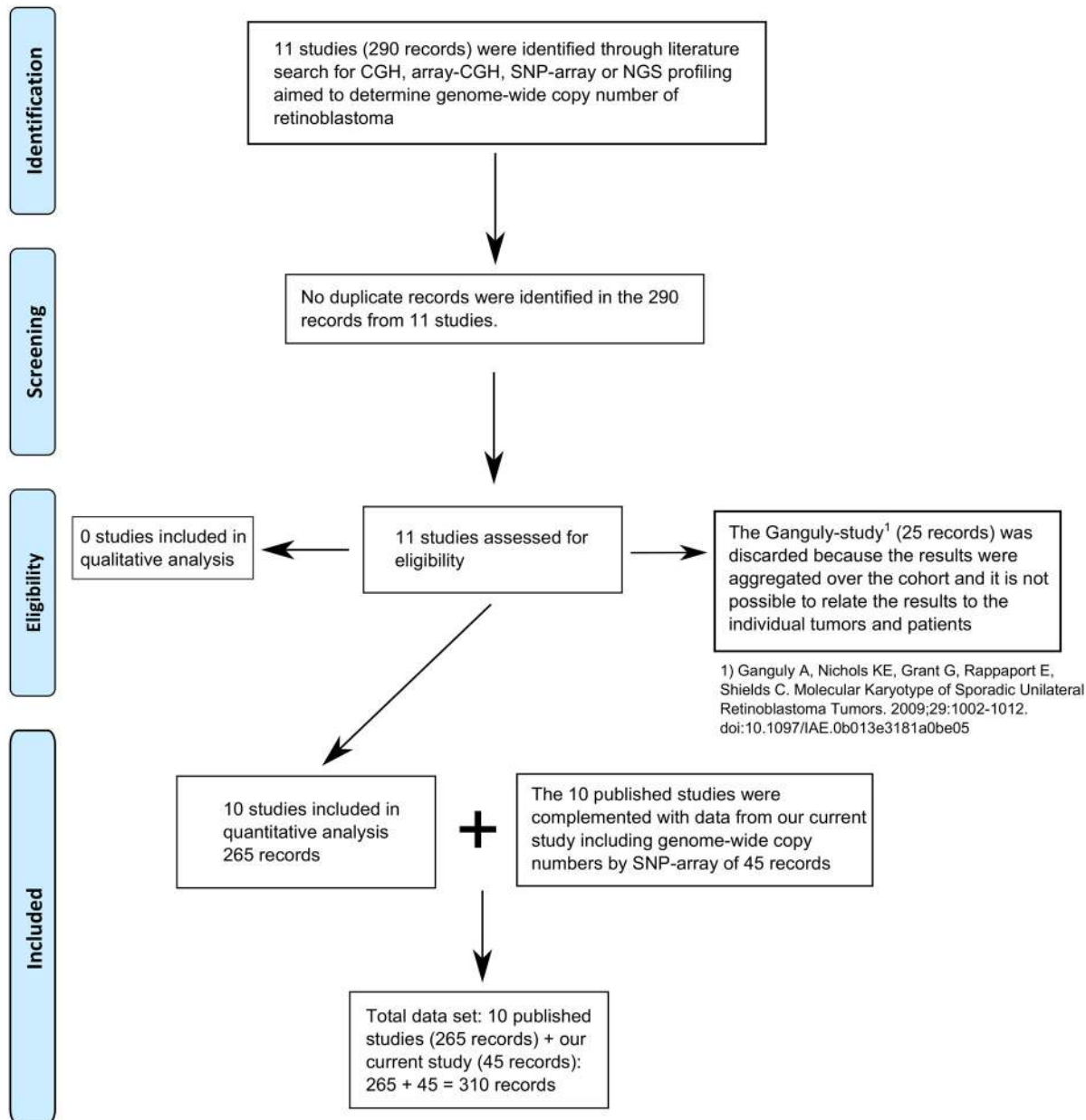
### Good agreement in SCNA-frequencies between 11 independent studies

Our study describes SCNAs of 45 primary retinoblastoma samples together with SCNAs from 10 published studies (S1 Table) adding up to 310 tumor samples (Fig 1, cohort description is given S2 Table). This allowed for a detailed genome-wide comparison of SCNAs determined from independent and international studies. For each HGNC approved gene, the percentage of the cohort affected by gains/losses is visualized along the genomic coordinates stratified by study (Fig 2). Percentages of SCNAs showed good agreement between studies, except for studies with small sample size and/or platform related differences. In the two small-sized studies (Gratias; N = 2, van der Wal, N = 13), SCNA percentages more easily reached high numbers. For example, in the Gratias study [13], gain of chromosome 1 and 6 was observed in 100% of the cohort, yet only affecting two patients. In addition, some platform specific differences between studies were reflected in the SCNA-percentages. For example, high-resolution studies (SNP-arrays, Kooi, Mol, Zhang) were able to detect small (< 100 Kb) SCNAs which resulted in more spikes in SCNA-frequencies. Three notable SCNA-gain spikes at 7p14.1, 7q34 and 14q11.2 overlapping with the TCR- $\gamma$ , TCR- $\beta$  and TCR- $\alpha/\delta$  gene clusters respectively, likely were not cancer related. It was shown that loss of these three regions is a frequent and somatic event that occurs in lymphocytes [18]. As a result, the three spikes of SCNA-gains appeared in the tumor-blood matched SNP-array analyses. Most likely these gains were the result of lymphocyte specific deletions and therefore the TCR- $\alpha/\beta/\gamma/\delta$  genes were omitted from further analyses in this study.

### Identification of candidate retinoblastoma driver genes

A major challenge in the interpretation of SCNAs is to distinguish driver from passenger genes since a single SCNA usually covers tens to hundreds of genes. To do so, SCNA-gain and SCNA-loss frequencies were used to compile a list of most plausible candidate genes and were integrated with micro-array gene expression data. For each approved HGNC gene, the number of patients with SCNA-gains subtracted from the number of patients with losses was calculated, which we call the SCNA-gain-loss difference. By subtracting the number of patients with gene losses from the number of patients with gene gains, non-disease associated SCNAs that arose from random genomic instability, which usually are both gained and lost in a cohort, were not prioritized. In addition to the SCNA-gain-loss difference, the percentage of patients affected by gain/loss of the respective gene (SCNA-percentage) was used as an extra threshold for candidate gene selection.

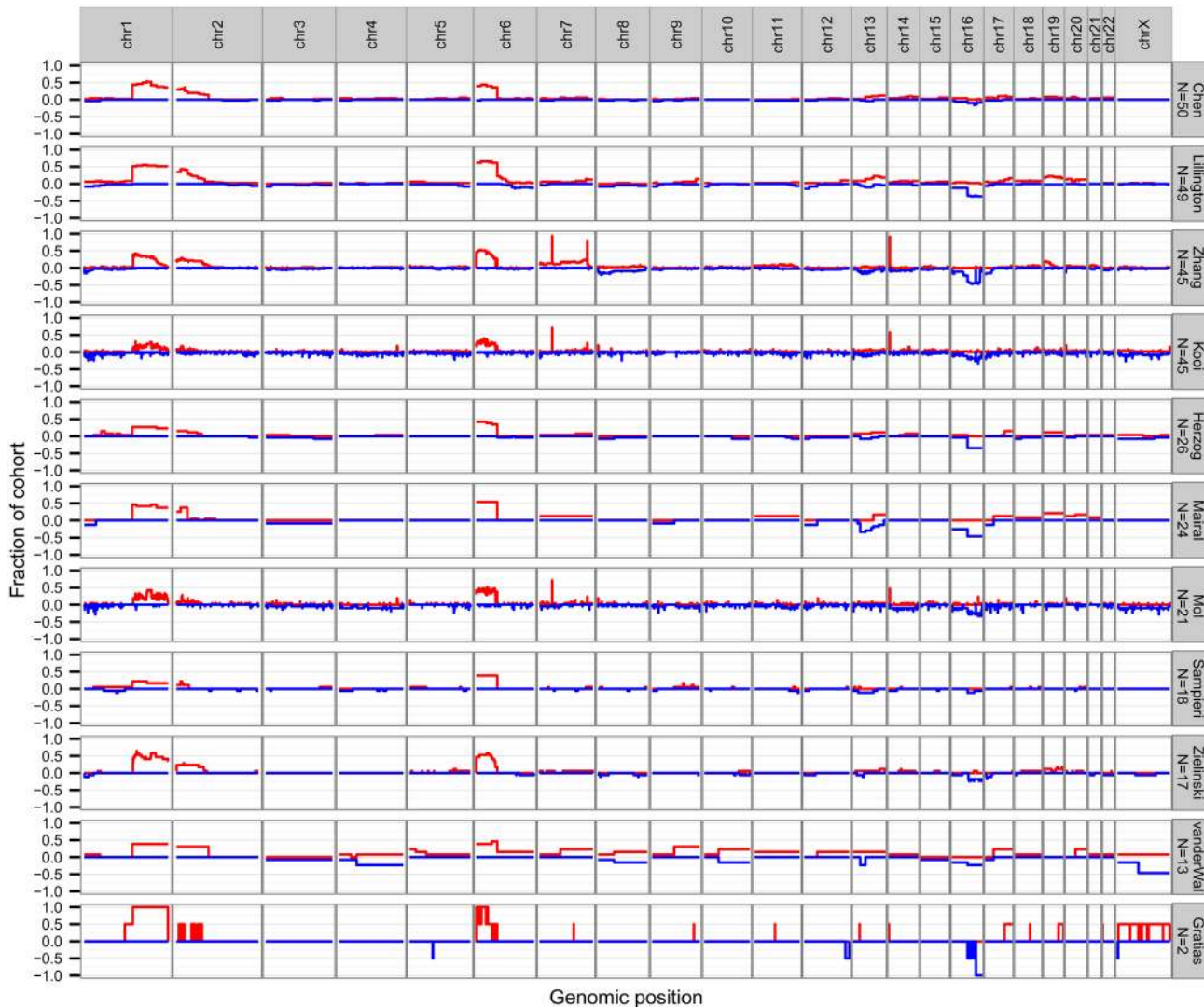
We considered genes with an SCNA-percentage >10% (out of 310 retinoblastoma samples) as candidate genes. This frequency threshold was empirically determined by plotting the number of genes that meet increasing SCNA percentage criteria (S1 Fig). This figure shows that the number of genes passing the SCNA percentage filter decreases rapidly but stabilizes at SCNA percentage between 5 and 15%. Chromosomes that contained candidate genes included chromosomes 1, 2, 6, 7, 13, 16 and 19. For these chromosomes, the SCNA-gain-loss difference is visualized along genomic coordinates (Fig 3). For each chromosome, peaks were defined by the gene with the highest SCNA-gain-loss difference and neighboring genes that showed at maximum 1% decrease in SCNA-gain-loss difference relative to the peak gene, visualized by the dashed numbered rectangles (Fig 3). Genomic coordinates with annotated gene symbols are given (S3 Table). For the Mol study [15] and our current study, matching gene expression profiling was available for 56 samples (S4 Table) [19]. Using this dataset, gene dosage effects (more/less gene copies is correlated with more/less gene expression) were quantified. For each



**Fig 1. PRISMA flow chart.** Searching google scholar for “retinoblastoma”, “copy number”, “(a)CGH”, and “SNP-array”, 11 studies were identified that performed genome-wide profiling of retinoblastoma adding up to 290 samples. No duplicate samples were identified. For the Ganguly study [14], copy number results could not be linked to individual tumors and therefore this study, including 25 samples was discarded. The remaining 265 samples were all included in quantitative analysis and were complemented with 45 SNP-arrays from our current study, adding up to 310 samples.

doi:10.1371/journal.pone.0153323.g001

peak region, genes with a significant gene dosage effect are listed in the last column of [S3 Table](#). Whereas the candidate list for chromosomes 1, 2, 6, 7 and 13 (relatively small peak regions) was narrowed down to a handful of genes, for chromosomes 16 and 19 (relatively large peak regions) there were still many candidate genes remaining. In [S5 Table](#), more detailed information is given on the candidate genes including the mean expression of the respective gene in the expression profiling cohort. For candidate genes on chromosome 1, 2, 6, and 7, genes with the highest mean expression were *NEK7*, *MYCN* and *DDX1*, *E2F3* and *SOX4*, and *CHCHD3*

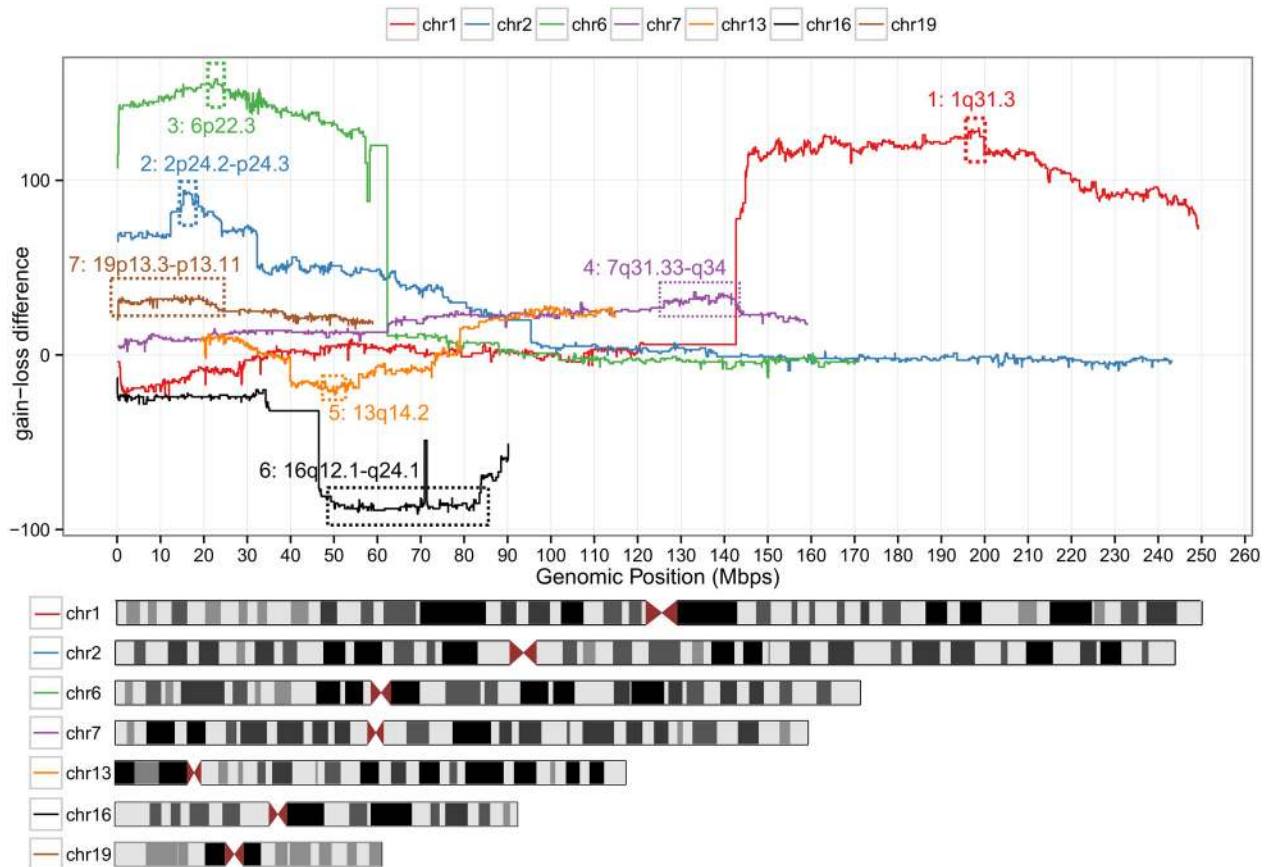


**Fig 2. SCNA frequencies per study.** Genome-wide SCNA-frequencies (DNA copy number gains in red, losses in blue) detected in primary retinoblastoma samples in our current study (Kooi, N = 45) and in ten published studies. Studies are ordered by sample size where the largest study is placed at the top.

doi:10.1371/journal.pone.0153323.g002

respectively. For chromosomes 13 containing deletions, *RBI* had the lowest mean expression. For candidates on chromosomes 16 and 19 there were no genes that clearly showed the lowest or highest expression. Therefore, narrowing down retinoblastoma-associated genes any further than presented in [S3 Table](#) was considered too speculative and was therefore omitted for chromosomes 16 and 19.

To test the robustness of the peaks identified by the described meta-analysis ([S3 Table](#)) and to identify smaller or additional peak regions, subset analyses were performed on the high-resolution SNP-array studies for which raw data was available (Mol, Zhang, Kooi, N = 111). In addition to DNACopy (circular binary segmentation), genoCN (Hidden Markov-Model) was used to infer the copy number states gain, loss and unchanged. The resulting gene-wise frequencies of gain and loss ([S2 Fig](#)) are visualized per study. The gain-loss differences determined by genoCN were strongly correlated (correlation test p-value < 2.2E-16, r = 0.95) with the gain-loss differences determined with DNACopy ([S5 Table](#)). Yet for 1q and 6p, the gain-loss



**Fig 3. Identification of SCNA peak regions.** For chromosomes that contained commonly altered genes, the gain-loss difference (number of patients with gains minus the number of patients with losses) is plotted for each official HGNC along chromosomal coordinates. For each of these chromosomes, peak regions were defined (also see [S3 Table](#)) indicated by the dashed rectangles and were used for retinoblastoma driver discovery.

doi:10.1371/journal.pone.0153323.g003

differences determined by genoCN reached a maximum at slightly different genes; for 1q at *MIR181* and for 6p at *KDM1B*, *DEK*, *RNA6-263P*, *RNF144B* and *MIR548A1* ([S5 Table](#)). The *DEK* gene displayed the most significant gene-dosage effect (FDR-adjusted p-value 3.25E-05) and is one of the most highly expressed genes interrogated by the micro-array (rank 38/18,290). In addition to genoCN segmentation, GISTIC analysis was performed on the Mol, Zhang and Kooi subset ([S3 Fig](#), [S5](#) and [S6](#) Tables). This algorithm uses not only SCNA frequency and recurrence but also the SCNA amplitude to identify significantly altered regions. GISTIC analysis confirmed the significance of arm-level gains at 1q and 6p and loss at 13q and 16q and also confirmed peak 2 (*MYCN*, 2p24.3), which was identified by the meta-analysis. While GISTIC analysis could not identify a clear focal peak at 6p, multiple peaks were identified at 1q and 13q, signifying the difficulty of single-gene identification. The highest peak at 1q in GISTIC analysis at 1q32.1 included 39 genes, but did not overlap with peak 1 identified by the meta-analysis ([S3 Table](#)). The GISTIC peaks at 13q also did neither overlap with corresponding peak 5 (13q) from the meta-analysis nor with the *RBI* locus. Additionally to the peaks identified by the meta-analysis ([S3 Table](#)), GISTIC identified focal gain of 14q22.3 including *OTX2* (gene dosage effect FDR-adjusted p-value 0.07) also discussed in the Mol study, loss of 1p36.32 (138 genes, [S5](#) and [S6](#) Tables) and loss of 17p13.3 (80 genes, [S5](#) and [S6](#) Tables). Of note, the GISTIC gain peaks at 7p14.1, 7q34 and 14q11.2 are the regions that are

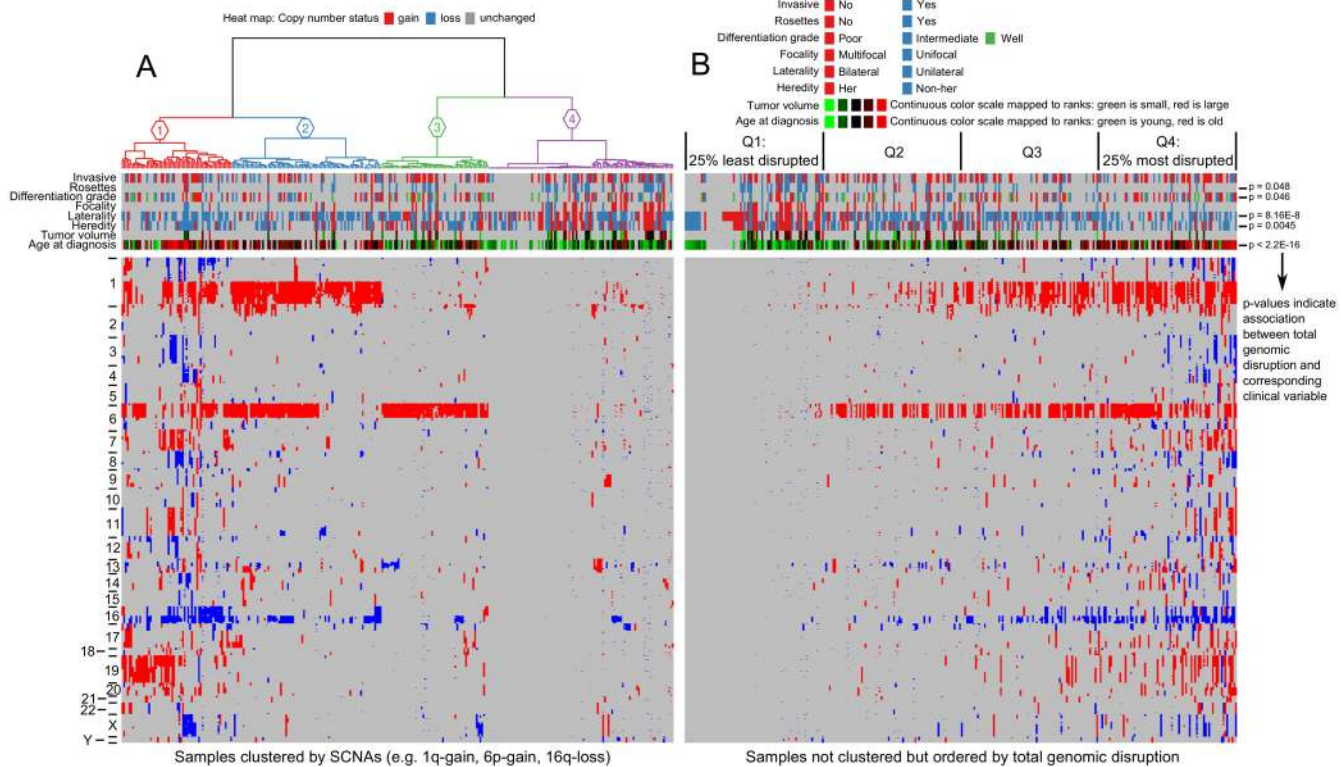
commonly lost in lymphocyte DNA, not gained by retinoblastomas (also see section 4.1, [Fig 2](#)) [[18](#)].

## Clustering of tumors based on SCNAs is mainly driven by total genomic disruption

Previous studies have shown that there is profound variability in copy number profiles of retinoblastoma samples associated significantly with age at diagnosis [[7,8,12,15](#)], heredity [[15](#)] and laterality [[12,15](#)]. In addition, profound differences in gene expression profiles between retinoblastoma samples were identified [[19,20](#)]. To identify genomic retinoblastoma subtypes, unsupervised hierarchical clustering (UHC) of retinoblastoma samples was performed using per gene SCNA-calls, displayed by a heat map ([Fig 4A](#)). The resulting retinoblastoma sample clustering is visualized by the dendrogram on top of the SCNA heat map together with corresponding color-coded sample information. The dendrogram was pruned to yield 4 UHC groups, optimizing for the differences between within cluster—and between cluster distances. Yet, it is arguable whether these 4 UHC groups represent truly distinct molecular retinoblastoma subtypes. To test for mutual exclusivity between frequently altered chromosomes, correlation tests were performed. None of the frequently altered chromosomes were significantly anti-correlated ([S4 Fig](#)). Instead of mutually exclusive SCNAs, clustering ([Fig 4A](#)) could be mainly driven by gradual variability in total genomic disruption. Indeed, total genomic disruption (in our study defined as the number of genes affected by SCNAs) was significantly different between the four UHC clusters. (Kruskal-Wallis p-value < 2.2E-16, Anova F-test: p-value < 2.2E-16). To further investigate a possible gradual variability in total genomic disruption, clustering ([Fig 4A](#)) was complemented by ordering the samples based on total genomic disruption ([Fig 4B](#)). There was a remarkable gradual increase in total genomic disruption that correlated with age at diagnosis (p-value < 2.2E-16). For each individual recurrent SCNA-gene, the frequency of occurrence increased with increasing total genomic disruption and age at diagnosis. Furthermore, recurrent 2p (mean age 25, standard error of the mean (SEM) 2.2 months) and 6p gains (mean 26, SEM 2.2 months) were observed in more stable and early diagnosed tumors (one-way ANOVA p-value 0.01) than 1q gains (mean 30 months, SEM 1.9 months) and 16q losses (mean 34, SEM 2.3 months). Possibly, total genomic disruption could be a better descriptive for SCNA-profiles than stratified groups identified by UHC.

## Genomic disruption increases with age at diagnosis, loss of differentiation, and SCNA-signal strength

To assess the statistical significance of the association between total genomic disruption and clinical and histopathological variables, hypothesis testing was performed and presented in [Table 1](#). Data of variables that significantly associated with total genomic disruption were also visualized ([Fig 5](#)). In case the clinical variable was numeric (e.g. age at diagnosis) the tumors were stratified in 4 disruption quartiles (Q1 = 25% least disrupted tumors, Q4 = 25% most disrupted tumors, [Fig 4B](#)) each showing boxplots of the clinical variable of interest. By definition, total genomic disruption differed significantly (p-value < 2.2e-16) between these disruption quartiles ([Fig 5A](#)). The average SCNA-amplitude per tumor was calculated from the segmentation mean of SCNAs with amplitudes both below and above the used segmentation threshold. Total genomic disruption linearly increased with SCNA-amplitudes ([Fig 5B](#)). This panel also shows that SCNA-amplitude rarely exceeds change of one copy (copy number 3, Log<sub>2</sub>-ratio 0.58). This means that there must be sample heterogeneity, either by intra-tumor clonal heterogeneity or contamination with non-cancer cells (e.g. retina or blood). For 40/66 (61%) samples in the Mol and our current study, DNA diagnostics identified LOH at the *RB1* allele as one of the disease causing



**Fig 4. Clustering of genome-wide SCNA.** Heat map of genome-wide SCNAs for 310 primary retinoblastoma samples (columns). Gains are indicated in red and losses in blue. (A) Samples are ordered by unsupervised hierarchical clustering (UHC) visualized by the dendrogram on top. The dendrogram was pruned in 4 UHC-groups indicated by UHC-group numbers and colored branches. Color-coded sample information: invasion: red = not-invasive, blue = invasive; rosettes: red = no rosettes, blue = rosettes; differentiation: red = poor, blue = moderate, green = well; focality: red = multifocal, blue = unifocal; laterality: red = bilateral, blue = unilateral; heredity: red = hereditary, blue = non-hereditary, tumor volume by MRI and age at diagnosis: continuous green-to-red scale where green means small tumors diagnosed at early age and vice versa. (B) Instead of ordering by UHC, samples are ordered based on total genomic disruption by SCNAs. P-values indicate the significance of association between total genomic disruption and the clinical variable.

doi:10.1371/journal.pone.0153323.g004

events (S7 Table). For these tumors, the homozygosity at the *RB1* locus (mBAF values, S7 Table) is indicative for tumor cellularity. Tumor purity was estimated to be very high (mBAF  $\geq 0.99$ ) for 22/40 (55%) tumors, high (mBAF  $\geq 0.90$ ) for 14/40 (35%) and moderate (mBAF  $\geq 0.74$ ) for 4/38 tumors. Also in the tumors for which the tumor cellularity was estimated to be very high, the SCNA amplitudes rarely exceeded one copy. An example is given for tumor 101032-02 (S5 Fig) which clearly shows LOH at 13q and incomplete LOH at 16q. Since the tumor cellularity for the rest of the cohort (272/310, 88%) could not be determined, possible non-cancer cell contamination could not be ruled out for the majority of the cohort.

Possibly, the gradual differences in SCNA-amplitudes are caused by differences in within-tumor heterogeneity. It has been demonstrated before that within the same tumor, fields of SCNA-devoid differentiated benign precursor lesions were located adjacent to more undifferentiated malignant retinoblastoma fields full of SCNAs [5]. In agreement, increase in SCNA-amplitudes and total genomic disruption correlated with decreased differentiation grades (p-value 0.04).

### Retinoblastoma cell lines showed high total genomic disruption

Cell lines derived from retinoblastoma primary tissue are considered valuable model systems to study retinoblastoma *in vitro*. To assess the genomic resemblance of retinoblastoma cell



**Table 1. Total genomic disruption in relation to clinical and histopathological variables.**

Independent variable	Descriptive statistics	Hypothesis testing p-value
<b>Age at diagnosis (months)</b>	<b>Mean ± SEM</b>	<2.20E-016* (WRST)
Q1 (25% least disrupted tumors)	8 ± 1.4 months	
Q2	14 ± 1.8 months	
Q3	26 ± 2.6 months	
Q4 (25% most disrupted tumors)	38 ± 3.4 months	
<b>Laterality</b>	<b># genes with SCNAs:</b>	8.16E-008* (WRST)
bilateral	1267 ± 230 genes	
unilateral	2845 ± 223 genes	
<b>Heredity</b>	<b># genes with SCNAs:</b>	0.0045* (WRST)
hereditary	1618 ± 252 genes	
non-hereditary	2752 ± 285 genes	
<b>Differentiation grade</b>	<b># genes with SCNAs:</b>	0.046* (LBL)
well	1998 ± 490 genes	
moderate	3307 ± 523 genes	
poor	3556 ± 420 genes	
<b>FW-rosettes</b>	<b># genes with SCNAs:</b>	0.048 (WRST)
yes	1293 ± 327 genes	
no	2661 ± 844 genes	
<b>Focality</b>	<b># genes with SCNAs:</b>	0.1144 (WRST)
multifocal	1016 ± 157 genes	
unifocal	2650 ± 658 genes	
<b>Volume measured by MRI (mm<sup>3</sup>)</b>	<b>Mean ± SEM</b>	0.1828 (WRST)
Q1 (25% least disrupted tumors)Q2	1028 ± 91 mm <sup>3</sup> 976 ± 123 mm <sup>3</sup>	
Q3	773 ± 153 mm <sup>3</sup>	
Q4 (25% most disrupted tumors)	1120 ± 173 mm <sup>3</sup>	
<b>Sex</b>	<b># genes with SCNAs:</b>	0.2191 (WRST)
male	2732 ± 278 genes	
female	2231 ± 228 genes	
<b>Rosettes</b>	<b># genes with SCNAs:</b>	0.4004 (WRST)
yes	1325 ± 225 genes	
no	2303 ± 986 genes	
<b>Familiarity</b>	<b># genes with SCNAs:</b>	0.4437 (WRST)
familial	1045 ± 651 genes	
non-familial	1507 ± 220 genes	
<b>Invasion</b>	<b># genes with SCNAs:</b>	0.5951 (WRST)
invasive	3094 ± 326 genes	
non-invasive	2939 ± 457 genes	

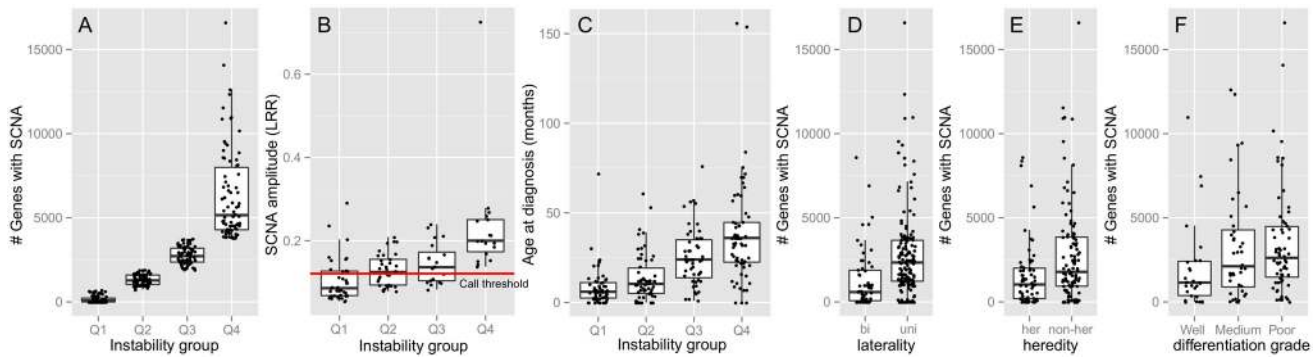
Tested variables are sorted by statistical significance. Q1-4 denote cohort quartiles based on the number of genes whose copy number is altered where Q1 denotes the cohort quartile with the least copy number altered genes and Q4 the cohort quartile with the most copy number altered genes.

WRST-Wilcoxon rank-sum test, LBL = Linear-by-linear.

\* indicates significance (p-value < 0.05). P-values lower than 0.001 are given in scientific notation.

doi:10.1371/journal.pone.0153323.t001

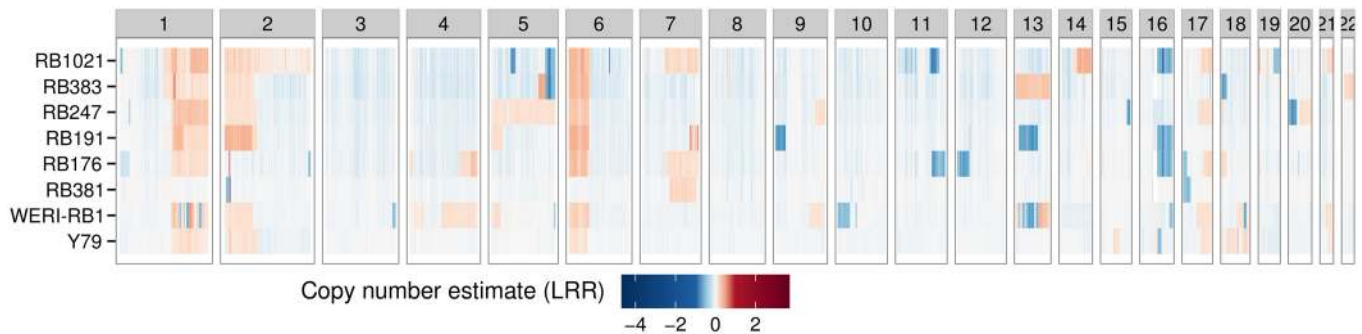
lines to primary tumors, we determined genome-wide SCNA-profiles for 8 retinoblastoma cell lines and compared those to the 45 primary retinoblastoma samples. Furthermore, retinoblastoma cell cultures have been extensively selected for proliferation by *in vitro* culturing and might reveal focal SCNAs driving retinoblastoma proliferation that remained undiscovered in



**Fig 5. Relation of total genomic disruption with clinical data.** Total genomic disruption by SCNAs with significantly correlated (*p*-values in Table 1) clinical and histopathological variables. Total genomic disruption is defined by the number of genes with either increased or decreased copy number. For numerical variables (B) age at diagnosis and (C) SCNA-amplitude, samples were categorized in quartiles based (A) on total genomic disruption. For example, Q1 denotes the 25% of tumors with the least genomic disruption and Q4 vice versa. SCNA-amplitudes are expressed as Log<sub>2</sub>-ratios. The horizontal red line illustrates the SCNA-threshold used for calling gains and losses (also see paragraph 2.4 in materials and methods and S6 Fig). For factorial variables (D) laterality, (E) heredity, and (F) differentiation grade the number of genes with SCNAs is stratified per factor level.

doi:10.1371/journal.pone.0153323.g005

primary samples. Segmented SCNAs are visualized in a heat map for autosomal chromosomes per cell line (Fig 6). Chromosomes 1, 2 and 6 were gained in 7/8 cell lines (RB1021, RB383, RB247, RB191, RB176, WERI-RB1, and Y79) and chromosome 16q was (partially) lost in 5/8 (63%) cell lines (RB1021, RB191 and RB176, WERI-RB1, and Y79). Gain of chromosome 7 was observed in 4/8 (50%) cell lines (RB1021, RB191, RB176 and RB381) whereas cell line RB191 showed focal increase of copy number at 7q33 (135–136.5 Mb). This region includes 11 genes (*CNOT4*, *SDHDP2*, *NUP205*, *C7orf73*, *RNU6-1154P*, *SLC13A4*, *FAM180A*, *MTPN*, *LUZP6*, *RNU6-223P*, and *PSMC1P3*). Only *NUP205* showed a significant association (FDR-adjusted *p*-value 2.62E-5) between copy numbers and expression and was also identified as a candidate in the primary tumor data set (S3 Table). Gain of chromosome 19q was not observed in any of the cell lines. The mean number of genes altered in cell lines (5813, S.E.M 1021) was significantly higher than in Q1 (*p*-value = 2.2E-06), Q2 (*p*-value = 3.7E-06) and Q3 (*p*-value = 1.3E-04), while not different from Q4 (*p*-value 0.65). This analysis indicates that retinoblastoma cell lines resemble primary tumors with high total genomic disruption or that they are a representation of the genomic disrupted part of the original tumors.



**Fig 6. Copy numbers in retinoblastoma cell lines.** Genome-wide heat map of SCNAs in 8 retinoblastoma cell lines. Sex chromosomes are not displayed since no matching reference was available. SCNA-amplitudes are expressed as Log<sub>2</sub>-ratio and mapped to a blue-to-red continuous color-scale.

doi:10.1371/journal.pone.0153323.g006

## Discussion

In our study, SCNA-profiles of 45 primary retinoblastoma samples were determined, which were analyzed together with SCNA profiles reported in ten published studies. In addition, the copy number data was integrated with publicly available matching gene expression data to further aid driver discovery. Candidate driver genes included *CRB1*, *NEK7* (1q), *MYCN* (2p), *SOX4* (6p), *RBI* (13q) and numerous genes for 16q. By dedicated subset analysis with the high-resolution platforms and by using alternative analysis algorithms, *MIR181* (1q) and *DEK* (6p) were identified additionally. Furthermore, our study shows examples of tumors with SCNAs do not exceed change of one copy, despite little non-cancer cell contamination, indicative for intra-tumor heterogeneity. Also, our meta-analysis allowed for a comprehensive association of retinoblastoma genotypes to the clinical phenotypes, which furthers our understanding of retinoblastoma.

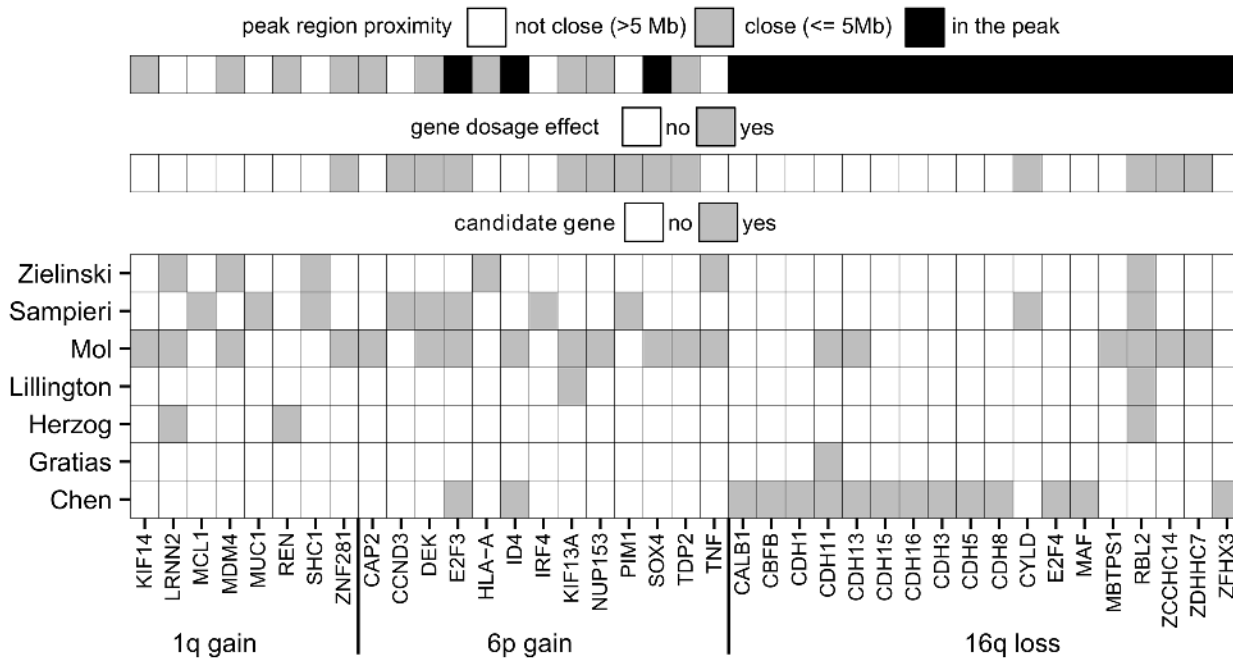
## Candidate genes reported in previous studies

Several studies aimed to identify genetic alterations promoting retinoblastoma development beyond loss of *RBI*. While most studies have focused on genetic alterations, it was also shown that epigenetic alterations might be important for retinoblastoma carcinogenesis [16]. Yet, the main focus of our current study is genetic alterations. Some of the previous retinoblastoma copy number alteration studies limited their discussion of SCNA-profiles to correlations with total genomic disruption (Mairal, van der Wal, Zhang), while other studies also provided suggestions for putative candidate genes beyond focally altered *MYCN* or *RBI* (Chen, Herzog, Lillington, Zielinski, Gratias, Sampieri, Mol). Although our study showed that the detected genome-wide SCNA-profiles showed good agreement between these studies (Fig 2), there was a clear variability in the suggested candidate genes between different studies (Fig 7).

## Candidate genes identified by the meta-analysis

For chromosome 1q-gains, eight different candidate genes have been proposed by previous studies. All of them are not located within the SCNA-gain candidate peak 1 region. Nevertheless, *KIF14*, *MDM4*, *REN*, and *ZNF281* were close (<5 Mb, see Fig 7, S5 Table) and *MDM4* and *REN* were located in the GISTIC peaks (S3 Fig, S5 and S6 Tables). In the meta-analysis, the most frequently gained genes with a significant gene dosage effect at chromosome 1q were *ZBTB41*, *CRB1* and *NEK7*. The *NEK7* gene showed the highest mean expression among these three candidates (S5 Table). The *NEK7* gene is part of the NIMA-related mitotic kinase gene family. In agreement, it was found that malignant retinoblastoma fields consist of mitotically active cells, in contrast to retinoma fields [21]. Also, overexpression of family member *NEK6* was shown to antagonize p53-induced senescence in human cancer cells [22]. Interestingly, benign precursor retinoma lesions stained positively for senescence-associated proteins p16<sup>INK4A</sup> and p130 [5]. Possibly, gain of *NEK7* and subsequent protein overexpression can antagonize p53-induced cellular senescence and causes benign retinoma cells to progress through the cell cycle. Furthermore, the oncogenic potential of *NEK6/7* has been recognized in various cancers, including breast cancer [23], gallbladder cancer [24], Wilm's tumors [25] and head and neck cancers [26]. Using our integrative genome-wide approach, our study now also identified *NEK7* as novel 1q candidate gene potentially driving retinoblastoma progression.

Next to *NEK7*, the second most highly expressed gene in peak 1, *CRB1*, is a noteworthy candidate driver gene as well. *CRB1* is involved in the development of photoreceptor cells [27], which are suggested to be the cells of origin of retinoblastoma [28]. Expression of *CRB1* interrupts naturally occurring apoptosis and photoreceptor apoptosis required for proper retinal morphogenesis [29]. Mutations in the *CRB1* gene cause abnormally thick retina with abnormal



**Fig 7. Reported candidate genes.** Overview of putative candidate genes proposed by previous studies for commonly gained chromosomal arms 1q and 6p and commonly lost 16q.

doi:10.1371/journal.pone.0153323.g007

lamination, in particular in the photoreceptor-dense are at the fovea [29]. Possibly, overexpression of *CRB1* driven by copy number gains accelerates photoreceptor-derived retinoblastoma.

Good concordance between our results and previously proposed candidate genes was observed for chromosome 6p gains. Out of 13 previously proposed candidates, *E2F3*, *ID4*, and *SOX4* are located in the candidate peak 2 region. For *ID4*, no significant gene dosage effect was observed in this study. To compensate for possible residual *RB1* activity, gain of *E2F3* might be important for retinoblastoma to develop. However, *SOX4* is also an interesting candidate gene. In hepatocellular carcinoma it was shown that *SOX4* over-expression led to a significant repression of p53-induced Bax expression and subsequent repression of p53-mediated apoptosis induced by gamma-irradiation [30]. Possibly, gain of *SOX4* in retinoblastoma could be a relevant hit beyond loss of *RB1*, allowing *RB1*-inactivated cells to better escape p53-induced apoptosis or senescence. Similarly to *NEK7*, *SOX4* has been identified as an important oncogene in a variety of other cancers including endometrial cancer [31], brain cancer [32,33], breast cancer [34], bladder cancer [35], ovarian cancer [36], colorectal cancer [37], liver cancer [30] and leukemia [38]. Therefore we suggest that *SOX4* should be considered as a serious candidate gene of the 6p gain region.

In case of chromosome 7q and 19p gains and 16q losses assignment of the driving gene is more speculative. Since focal SCNAs are not observed at these genomic loci, the resulting candidate peak regions contain numerous candidate genes. Several previous studies proposed *RBL2* (protein p130) as a candidate gene for the 16q loss (Fig 7) which is often seen in patients diagnosed at later age. Since p130 is primarily expressed in  $G_0$ -cells restricting them for cell cycle entry [39], loss of p130 could prevent cellular senescence and promote the benign-to-malignant transition. Our study showed that copy numbers of *RBL2* are commonly reduced and are associated with decreased expression. Therefore, also in our analysis, it remains one of the many candidates for 16q loss. For chromosome 7q gains, out of 10 proposed candidates

based on primary retinoblastoma samples, only *NUP205* is included in a focal gain observed in cell line RB191. In lung cancer cell lines it was shown that through TMEM209 stabilization of *NUP205*, protein levels of *MYC* were increased and promoted cell growth. Attenuation of TMEM209 stabilization corresponded to blocked growth, indicating the TMEM209-NUP250 complex might play a role in cell proliferation [40].

### Limitations of candidate gene identification

By integrating data from our DNA profiling study with published studies and integration with gene expression data, we present a comprehensive effort to identify retinoblastoma driving genes. A potential disadvantage of data pooling is that the resolution of the pooled data is lower than the resolution of the three high-resolution SNP-array studies. Another disadvantage is that SCNA amplitudes are not available for all studies. This meant that in the pooled analyses, SCNA amplitudes were not taken into account. Therefore, the pooled analysis was complemented by a subset GISTIC analysis using the high-resolution Mol, Zhang and Kooi datasets only (S3 Fig, S5 and S6 Tables). While GISTIC analysis confirmed the significance of 1q, 6p and 16q alterations and peak 2 (2p24.3, including *MYCN*) from the pooled analysis, the peak regions at 1q and 13q were slightly shifted in GISTIC analysis. For 1q, GISTIC even identified multiple regions to be significantly altered. Also when a different segmentation algorithm was used (genoCN, S2 Fig and S5 Table), the peaks at 1q and 6p slightly shifted, in this case towards the *MIR181* and *DEK* genes respectively. Except for 2p where only *MYCN* is included in all gained regions, multiple candidate genes remain for the commonly altered chromosomes. Therefore, functional assays are needed to empirically determine the oncogenic potential of the described candidate genes.

### Models explaining the association between total genomic disruption and age at diagnosis

Previous studies showed that retinoblastoma tumors have profound variability in total genomic disruption. It was unclear whether this variability is dichotomous or gradual, suggestive of two subtypes or gradual progression respectively. The study of van der Wal et al. suggested that variability in total genomic disruption was bimodal, although this was based on a small study ( $N = 13$ ) and was not substantiated by any statistics. In the study of Mol et al. ( $N = 21$ ), it was shown that unsupervised hierarchical clustering divided retinoblastoma samples into three branches with increasing total genomic disruption. Our study conclusively shows that total genomic disruption is gradual and co-occurrence and/or mutual exclusivity of SCNAs is not apparent. Increasing total genomic disruption was related to decreasing differentiation grades and suggests a de-differentiation process. Since our data includes indications for intra-tumor heterogeneity of genomic alterations, possibly there was also heterogeneity in differentiation grades between cells within tumors. In agreement, examples have been described where fields of differentiated cells lie adjacent to undifferentiated cells [21].

It is interesting why tumors with much genomic disruption and poorly differentiated cells were particularly observed in patients diagnosed at late age. Possibly, in tumors where the second *RBI* hit occurred at later age the resulting precursor lesion was less proliferative than lesions that developed more early in retina development. When SCNAs occurs in these late-onset precancerous lesions, the initial SCNA-devoid cells are easily overgrown by the progressed proliferative cells. The hypothesis that the proliferative consequence of *RBI* inactivation in the retina is age-dependent is underscored by the fact that retinoblastoma does not occur after the retina is fully developed. Additionally, diagnosis could have been delayed in

patients with older age at diagnosis and thereby allowed the tumors more time to acquire SCNAs and progress.

## In conclusion

Our integrated approach allowed us to refine and improve the lists of putative retinoblastoma driving genes. This limited set of genes can serve as leads for future studies on retinoblastoma progression and precision medicine. However, we also found that for at least a subset of tumors, abnormal gene copy numbers were not always present in all tumor cells. Therefore, a multi-target treatment strategy might be required for efficient retinoblastoma treatment.

## Materials and Methods

### Tissue collection

Tumor samples were obtained from retinoblastoma patients after primary enucleation and peripheral blood samples were collected at initial presentation before treatment. In the Netherlands, all patients are referred to the VU University Medical Center. Hence a well-documented cohort of unselected primary enucleated eyes was available for molecular studies. Tumor samples were snap frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until further analysis. All patient samples and clinical and histopathological features were collected and stored according to local ethical regulations. All patients gave consent verbally, as this was the standard in the time the included patients were diagnosed. Since genetic analyses of our study focused on tumor DNA and not germ line DNA, waiver of informed consent was specifically given for genetic analyses by The Medical Ethics Review Committee of the VU University Medical Center which is registered with the USA OHRP as IRB00002991. The FWA number assigned to VU University Medical Center is FWA00017598. A cohort description including clinical and histopathological information is given in [S2 Table](#). RB cell lines RB1021, RB383, and RB247 [41] were kindly provided by the laboratory of Brenda Gallie and cell lines RB191, RB176, and RB381 [42] by the laboratory of David Cobrinik.

### DNA isolation

Genomic DNA from frozen tumor retinoblastoma specimens was isolated with the NucleoSpin Tissue kit (Macherey-Nagel, Düren, Germany) or Wizard Genomic DNA Purification Kit (Promega, Madison, USA). DNA quality was analyzed for high molecular bands  $>20$  Kb by agarose gel electrophoresis. DNA concentration and OD 260/280 ratio was determined with the Nanodrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, USA). DNA yields and quality were within the same range for all samples.

### SNP-arrays

Microarray-based DNA genotyping experiments were performed at ServiceXS (ServiceXS B.V., Leiden, The Netherlands) using the HumanOmni1-Quad BeadChip (Illumina, San Diego, USA), according to the manufacturer's instructions. The BeadChip images were scanned on the iScan system and the data was extracted into Illumina's GenomeStudio software v2010.1. The software's default settings were used with the cluster file as developed by Illumina for genotype calling. Resulting copy number estimates ( $\text{Log}_2$ -ratios between tumor and matched blood) and B Allele frequencies were normalized with tQn normalization [43] and segmented with DNACopy [44] with a minimum segment length of five markers. For loss of heterozygosity detection, segmentation of converted B allele frequencies (mBAF) was performed using BAF-segmentation using a mBAF threshold of 0.8 [45]. A minimum of five consecutive markers was

used for segmentation together with a minimum mBAF-amplitude of 0.6. In parallel to DNA-copy segmentation, genoCN segmentation was used with default parameters to infer copy number states gain, loss and unchanged of the SNP-array datasets. To identify significantly altered regions, GISTIC analysis was performed ( $q$ -value  $< 0.05$ ) using the combined segmentation (by DNACopy) results of the Mol, Zhang and Kooi data sets. Gene expression data is available at GSE59983 (primary samples) and GSE77094 (cell lines). DNA copy number data is available at EGAS00001001715.

## Data collection and analysis

By manual google scholar search, studies profiling SCNAs by CGH, array-CGH, SNP-array or NGS were identified (Fig 1). Studies that reported SCNAs by cytoband location were digitalized by manually looking up the current genomic coordinates of the reported cytobands (hg19). Careful examination of sample identifiers was performed to prevent duplicate records, yet no duplicate records were identified. In case SCNAs were reported in genomic coordinates, they were converted to hg19 coordinates by the UCSC liftover tool [46]. For studies where raw data was available [13,15,16], copy number estimates were segmented with DNACopy as described above. Since biopsies are uncommon in retinoblastoma, the included samples are possibly biased for later staged tumors that had to be enucleated. Furthermore, earlier studies used lower resolution profiling compared to more recent studies and therefore more subtle genomic alterations might have been more readily identified by later studies.

All published SCNAs were concatenated with SCNAs detected by our study. SCNAs were called in three states; loss, normal and gain using copy number thresholds 1.8 ( $\text{Log}_2$ -ratio = -0.15) for losses and 2.2 ( $\text{Log}_2$ -ratio = 0.14) for gains for segments with a  $p$ -value  $< 0.05$  (S6 Fig). For each official HGNC gene, the segmentation mean was calculated by overlapping genomic coordinates of the genes with detected SCNAs using BEDOPS [47]. For hierarchical clustering, Ward's agglomerative clustering was performed using Euclidean distances. Statistical analysis and visualization was done in R (Pumpkin Helmet, version 3.1.2). For hypothesis testing where both independent and dependent variables are numeric, Wilcoxon signed-rank tests were used. For 2-level categorical independent variables and numeric dependent variables, Wilcoxon rank-sum tests were used. For independent categorical variables with more than 2 unordered levels and numeric dependent variables, Kruskal-Wallis tests were used. For independent categorical variables with more than 2 ordered levels (e.g. differentiation grade low  $<$  medium  $<$  high) and dependent numerical variables, linear-by-linear association tests were used implemented by the "coin" R-package. Two-sided  $p$ -values below 0.05 were considered statistically significant. Extreme  $p$ -values lower than  $2.2\text{E-}16$  could not be calculated and are reported as  $<2.2\text{E-}16$ .

For the determination of gene-dosage effects, linear regression was performed of continuous copy number estimates (segmented  $\text{Log}_2$ -ratios) on RMA-normalized expression estimates. The linear regression slope was tested for significance, and Benjamini-Hochberg multiple-testing corrected  $p$ -values were calculated. Genes with a positive regression slope (the more DNA copies, the more gene expression) and multiple-testing corrected  $p$ -values  $< 0.05$  were considered to display a gene-dosage effect.

## Supporting Information

**S1 Fig. Determination of the SCNA percentage threshold.** The number of genes remaining (Y-axis) after applying more stringent criteria for common events (X-axis, SCNA percentage threshold) rapidly decreases and reaches a transition point at SCNA percentage threshold

between 5 and 10%. Our study defined SCNAs occurring in  $> 10\%$  of the cohort to be frequent. (TIFF)

**S2 Fig. Effect of Hidden Markov-Model segmentation.** To determine the robustness of DNA-copy segmentation (Fig 2), genoCN segmentation implementing a Hidden Markov-Model approach was applied to the Mol, Zhang and Kooi studies ( $N = 111$ ). For each gene, the frequency of gain and loss is given in the subset cohort. (TIFF)

**S3 Fig. Subset analysis with GISTIC.** Results of GISTIC analysis on the Mol, Zhang and Kooi dataset segmented by DNACopy ( $N = 111$ ). Regions with q-values below 0.05 were considered significant and are annotated with cytoband labels. Region coordinates and HGNC gene symbol annotations are provided in S6 Table. (TIFF)

**S4 Fig. Correlations between SCNA peaks.** Pearson correlation matrix testing for co-occurrence and mutual exclusivity between peak regions containing retinoblastoma-driving candidate genes. The lower-left triangle is a color-coded (blue = mutual exclusivity, red = co-occurrence) representation of the upper-right triangle which gives the Pearson correlations. Peak regions showed no mutual exclusivity and weak co-occurrence. The best correlation (0.42) was found between 1q gain and 16q loss, both events often observed in patients diagnosed at late age. (TIFF)

**S5 Fig. Within-tumor heterogeneity of 16q-loss.** An example of a tumor sample without non-cancer cell contamination (100% LOH at *RBI*, chromosome 13), but with incomplete LOH of 16q. Only SNPs that were heterozygous in the matching blood sample were used for this analysis. (A) Overview of mirrored B-allele frequencies (mBAF) segmented with BAFsegmentation. This sample displayed 100% LOH of 13q illustrated by  $mBAF \sim 1$  indicating that this sample does not contain any detectable amounts of non-cancer cells. On the contrary, mBAF of 16q was segmented at mBAF 0.65, indicating that this sample contained cells with 16q-LOH (mBAF 1) and cells without 16q-LOH (mBAF 0.5). (B) B-allele frequencies of SNPs that were heterozygous in the matched germ line sample of chromosome 13 (complete LOH) and 16 (mixture of LOH and normal). Note that no data is available for the 13p-region since the DNA sequence of this region remains to be determined. (TIFF)

**S6 Fig. Determination of gain and loss thresholds.** Histograms of segmented Log2-ratios for Mol and Kooi (Illumina platform) and Zhang (Affymetrix platform) datasets. The dashed red and blue lines indicate the thresholds used for gains and losses respectively. (TIFF)

**S1 Table. Description of SCNA studies.** For each study included in the meta-analysis it is described how SCNAs were determined and integrated in our meta-analysis. (XLSX)

**S2 Table. Cohort description.** Statistics about the number of included samples per platform and patient phenotype variables. (XLSX)

**S3 Table. Determination of candidate driver regions by meta-analysis.** For chromosomes that were altered in at least 10% (31/310) of the pooled cohort, peak regions were identified. The peak regions were defined by the maximum SCNA gain-loss index for that chromosome



with peak boundaries at 1% deflection from the peak. The peak regions were annotated with genes that displayed a gene-dosage effect and were considered candidate genes driving retinoblastoma oncogenesis.

(XLSX)

**S4 Table. Copy number and expression values.** For 56 samples from the Mol and Kooi studies, both SNP-array and gene expression profiling was performed. For each gene that was profiled by both methods, the Log<sub>2</sub>-ratio (DNA) and Log<sub>2</sub>-transformed normalized expression (RNA) is given.

(XLSX)

**S5 Table. Per-gene SCNA-gain-loss values.** For each approved HGNC gene, the SCNA-gain-loss difference is given together with gene dosage effect testing. In a subset analysis on the Mol, Zhang and Kooi dataset, the gain-loss difference was also determined using Hidden Markov-Model based segmentation (genoCN). Additionally, GISTIC analysis was performed in this subset and the per-gene q-values are given, indicating the per-gene significance of alterations.

(XLSX)

**S6 Table. GISTIC subset analysis results.** Regions that were determined to be significantly (S3 Fig, q-value < 0.05) altered based on recurrence, amplitude and focality are given. Regions are annotated with HGNC gene symbols.

(XLSX)

**S7 Table. B-allele frequencies of *RB1*.** For the Mol study and the current study, SNP-arrays were used and conventional DNA diagnostics for *RB1* was available. Using this data, the tumor cellularity can be estimated. In case the second hit was LOH (40 tumors), B allele frequencies of the *RB1* allele should often (34/40 tumors) exceeded 0.95 indicating non-cancer contamination at maximum was 10% for these tumors.

(XLSX)

## Acknowledgments

We would like to thank all members of the VU University Medical Center Retinoblastoma team for feedback and discussions.

## Author Contributions

Conceived and designed the experiments: JCD JC HtR ACM GJLK HM-H. Performed the experiments: IK BMM. Analyzed the data: IK BMM MM. Wrote the paper: IK BM ACM HtR JC JCD HM-H GJLK. Provided pathological data: PvdV. Provided radiological data: McdJ PdG.

## References

1. Linabery AM, Ross JA. Trends in childhood cancer incidence in the U.S. (1992–2004). *Cancer*. 2008; 112: 416–32. doi: [10.1002/cncr.23169](https://doi.org/10.1002/cncr.23169) PMID: [18074355](https://pubmed.ncbi.nlm.nih.gov/18074355/)
2. Parkin DM, Stiller CA, Draper GJ, Bieber CA. The international incidence of childhood cancer. *Int J Cancer*. 1988; 42: 511–520. doi: [10.1002/ijc.2910420408](https://doi.org/10.1002/ijc.2910420408) PMID: [3170025](https://pubmed.ncbi.nlm.nih.gov/3170025/)
3. Knudson AG. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A*. 1971; 68: 820–823. PMID: [5279523](https://pubmed.ncbi.nlm.nih.gov/5279523/)
4. Rushlow DE, Mol BM, Kennett JY, Yee S, Pajovic S, Thériault BL, et al. Characterisation of retinoblastomas without RB1 mutations: genomic, gene expression, and clinical studies. *Lancet Oncol*. 2013; 14: 327–34. doi: [10.1016/S1470-2045\(13\)70045-7](https://doi.org/10.1016/S1470-2045(13)70045-7) PMID: [23498719](https://pubmed.ncbi.nlm.nih.gov/23498719/)

5. Dimaras H, Khetan V, Halliday W, Orlic M, Prigoda NL, Piovesan B, et al. Loss of RB1 induces non-proliferative retinoma: increasing genomic instability correlates with progression to retinoblastoma. *Hum Mol Genet.* 2008; 17: 1363–1372. doi: [10.1093/hmg/ddn024](https://doi.org/10.1093/hmg/ddn024) PMID: [18211953](https://pubmed.ncbi.nlm.nih.gov/18211953/)
6. Chen D, Gallie BL, Squire JA. Minimal regions of chromosomal imbalance in retinoblastoma detected by comparative genomic hybridization. *Cancer Genet. Division of Cancer Informatics, Ontario Cancer Institute/Princess Margaret Hospital, University Health Network, 610 University Avenue, M5G 2M9, Toronto, Canada; 2001; 129: 57–63.*
7. Herzog S, Lohmann DR, Buiting K, Schuler A, Horsthemke B, Rehder H, et al. Marked differences in unilateral isolated retinoblastomas from young and older children studied by comparative genomic hybridization. *HumGenet. Institut für Klinische Genetik, Zentrum für Humangenetik, Philipps-Universität, Marburg, Germany; 2001; 108: 98–104.*
8. Lillington DM, Kingston JE, Coen PG, Price E, Hungerford J, Domizio P, et al. Comparative genomic hybridization of 49 primary retinoblastoma tumors identifies chromosomal regions associated with histopathology, progression, and patient outcome. *Genes Chromosom. Department of Medical Oncology, St Bartholomew's Hospital Medical College and the Royal London NHS Trust, London, United Kingdom. D.Lillington@cancer.org.uk; 2003; 36: 121–128. Available: file:///g:/literature study Rb/CGH/lillington 2003 CGH of 49 RB tumors.pdf*
9. Mairal A, Pinglier E, Gilbert E, Peter M, Validire P, Desjardins L, et al. Detection of chromosome imbalances in retinoblastoma by parallel karyotype and CGH analyses. *Genes Chromosom. INSERM U509 Pathologie Moléculaire des Cancers, Institut Curie, Section de Recherche, Paris, France; 2000; 28: 370–379. Available: file:///g:/literature study Rb/CGH/mairal 2000 genes chromosomes cancer detection of chromosome imbalances in rb by parallel karyotype and CGH analyses.pdf*
10. van der Wal JE, Hermsen MAJA, Gille HJP, Schouten-van Meeteren NY., Moll A., Imhof SM, et al. Comparative genomic hybridisation divides retinoblastomas into a high and a low level chromosomal instability group. *J Clin Pathol.* 2003; 56: 26–31. PMID: [12499428](https://pubmed.ncbi.nlm.nih.gov/12499428/)
11. Zielinski B, Gratias S, Toedt G, Mendrzyk F, Stange DE, Radlwimmer B, et al. Detection of chromosomal imbalances in retinoblastoma by matrix-based comparative genomic hybridization. *Genes Chromosomes Cancer. German Cancer Research Center, Division of Molecular Genetics, Heidelberg, Germany; 2005; 43: 294–301. doi: 10.1002/gcc.20186*
12. Sampieri K, Amenduni M, Papa FT, Katzaki E, Mencarelli MA, Marozza A, et al. Array comparative genomic hybridization in retinoma and retinoblastoma tissues. *Cancer Sci.* 2009; 100: 465–471. doi: [10.1111/j.1349-7006.2008.01070.x](https://doi.org/10.1111/j.1349-7006.2008.01070.x) PMID: [19183342](https://pubmed.ncbi.nlm.nih.gov/19183342/)
13. Gratias S, Rieder H, Ullmann R, Klein-Hitpass L, Schneider S, Bölöni R, et al. Allelic loss in a minimal region on chromosome 16q24 is associated with vitreous seeding of retinoblastoma. *Cancer Res.* 2007; 67: 408–16. doi: [10.1158/0008-5472.CAN-06-1317](https://doi.org/10.1158/0008-5472.CAN-06-1317) PMID: [17210724](https://pubmed.ncbi.nlm.nih.gov/17210724/)
14. Ganguly A, Nichols KE, Grant G, Rappaport E, Shields C. Molecular karyotype of sporadic unilateral retinoblastoma tumors. *Retina.* 2009; 29: 1002–1012. PMID: [19491728](https://pubmed.ncbi.nlm.nih.gov/19491728/)
15. Mol BM, Massink MPG, van der Hout AH, Dommering CJ, Zaman JMA, Bosscha MI, et al. High resolution SNP array profiling identifies variability in retinoblastoma genome stability. *Genes Chromosomes Cancer.* 2014; 53: 1–14. doi: [10.1002/gcc.22111](https://doi.org/10.1002/gcc.22111) PMID: [24249257](https://pubmed.ncbi.nlm.nih.gov/24249257/)
16. Zhang J, Benavente C a., McEvoy J, Flores-Otero J, Ding L, Chen X, et al. A novel retinoblastoma therapy from genomic and epigenetic analyses. *Nature. Nature Publishing Group; 2012; 481: 329–34. doi: 10.1038/nature10733*
17. Sampieri K, Mencarelli MA, Epistolato MC, Toti P, Lazzi S, Bruttini M, et al. Genomic differences between retinoma and retinoblastoma. *Acta Oncol.* 2008; 47: 1483–1492. doi: [10.1080/02841860802342382](https://doi.org/10.1080/02841860802342382) PMID: [18785023](https://pubmed.ncbi.nlm.nih.gov/18785023/)
18. Schwienbacher C, De Grandi A, Fuchsberger C, Facheris MF, Svaldi M, Wjst M, et al. Copy number variation and association over T-cell receptor genes—influence of DNA source. *Immunogenetics.* 2010; 62: 561–7. doi: [10.1007/s00251-010-0459-7](https://doi.org/10.1007/s00251-010-0459-7) PMID: [20582410](https://pubmed.ncbi.nlm.nih.gov/20582410/)
19. Kooi IE, Mol BM, Moll AC, van der Valk P, de Jong MC, de Graaf P, et al. Loss of photoreceptor genes and gain of genomic alterations in retinoblastoma reveal tumor progression. *EBioMedicine.* 2015; doi: [10.1016/j.ebiom.2015.06.022](https://doi.org/10.1016/j.ebiom.2015.06.022)
20. Kapatai G, Brundler M-A, Jenkinson H, Kearns P, Parulekar M, Peet AC, et al. Gene expression profiling identifies different sub-types of retinoblastoma. *Br J Cancer. Cancer Research UK; 2013; 109: 512–25. doi: 10.1038/bjc.2013.283*
21. Eagle RC. High-Risk Features and Tumor Differentiation in Retinoblastoma. *Arch Pathol Lab Med.* 2009; 133: 1203–1209. Available: <http://www.archivesofpathology.org/doi/pdf/10.1043/1543-2165-133.8.1203> PMID: [19653710](https://pubmed.ncbi.nlm.nih.gov/19653710/)

22. Jee HJ, Kim AJ, Song N, Kim H-J, Kim M, Koh H, et al. Nek6 overexpression antagonizes p53-induced senescence in human cancer cells. *Cell Cycle*. 2010; 9: 4703–10. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21099361> PMID: [21099361](https://pubmed.ncbi.nlm.nih.gov/21099361/)
23. Nassirpour R, Shao L, Flanagan P, Abrams T, Jallal B, Smeal T, et al. Nek6 mediates human cancer cell transformation and is a potential cancer therapeutic target. *Mol Cancer Res*. 2010; 8: 717–28. doi: [10.1158/1541-7786.MCR-09-0291](https://doi.org/10.1158/1541-7786.MCR-09-0291) PMID: [20407017](https://pubmed.ncbi.nlm.nih.gov/20407017/)
24. Wang R, Song Y, Xu X, Wu Q, Liu C. The expression of Nek7, FoxM1, and Plk1 in gallbladder cancer and their relationships to clinicopathologic features and survival. *Clin Transl Oncol*. 2013; 15: 626–32. doi: [10.1007/s12094-012-0978-9](https://doi.org/10.1007/s12094-012-0978-9) PMID: [23359173](https://pubmed.ncbi.nlm.nih.gov/23359173/)
25. Natrajan R, Williams R, Fenwick K, Iravani M, Mackay A, Jones C, et al. Molecular profiling of relapsing Wilms' tumors using microarray-CGH. *AAO Meeting Abstr*. 2004; 2004: 627. Available: <http://www.aacrmeetingabstracts.org/cgi/content/abstract/2004/1/627-c>
26. Saloura V, Cho H-S, Kyiotani K, Alachkar H, Zuo Z, Nakakido M, et al. WHSC1 Promotes Oncogenesis through Regulation of NIMA-related-kinase-7 in Squamous Cell Carcinoma of the Head and Neck. *Mol Cancer Res*. 2014; 1541–7786.MCR-14-0292-T-. doi: [10.1158/1541-7786.MCR-14-0292-T](https://doi.org/10.1158/1541-7786.MCR-14-0292-T)
27. Mehalow AK, Kameya S, Smith RS, Hawes NL, Denegre JM, Young JA, et al. CRB1 is essential for external limiting membrane integrity and photoreceptor morphogenesis in the mammalian retina. *Hum Mol Genet*. 2003; 12: 2179–89. doi: [10.1093/hmg/ddg232](https://doi.org/10.1093/hmg/ddg232) PMID: [12915475](https://pubmed.ncbi.nlm.nih.gov/12915475/)
28. Xu XL, Singh HP, Wang L, Qi D-L, Poulos BK, Abramson DH, et al. Rb suppresses human cone-precursor-derived retinoblastoma tumours. *Nature*. 2014; 514: 385–388. doi: [10.1038/nature13813](https://doi.org/10.1038/nature13813) PMID: [25252974](https://pubmed.ncbi.nlm.nih.gov/25252974/)
29. Jacobson SG. Crumbs homolog 1 (CRB1) mutations result in a thick human retina with abnormal lamination. *Hum Mol Genet*. 2003; 12: 1073–1078. doi: [10.1093/hmg/ddg117](https://doi.org/10.1093/hmg/ddg117) PMID: [12700176](https://pubmed.ncbi.nlm.nih.gov/12700176/)
30. Hur W, Rhim H, Jung CK, Kim JD, Bae SH, Jang JW, et al. SOX4 overexpression regulates the p53-mediated apoptosis in hepatocellular carcinoma: clinical implication and functional analysis in vitro. *Carcinogenesis*. 2010; 31: 1298–307. doi: [10.1093/carcin/bgg072](https://doi.org/10.1093/carcin/bgg072) PMID: [20400479](https://pubmed.ncbi.nlm.nih.gov/20400479/)
31. Huang Y-W, Liu JC, Deatherage DE, Luo J, Mutch DG, Goodfellow PJ, et al. Epigenetic repression of microRNA-129-2 leads to overexpression of SOX4 oncogene in endometrial cancer. *Cancer Res*. 2009; 69: 9038–46. doi: [10.1158/0008-5472.CAN-09-1499](https://doi.org/10.1158/0008-5472.CAN-09-1499) PMID: [19887623](https://pubmed.ncbi.nlm.nih.gov/19887623/)
32. Lin B, Madan A, Yoon J-G, Fang X, Yan X, Kim T-K, et al. Massively parallel signature sequencing and bioinformatics analysis identifies up-regulation of TGFBI and SOX4 in human glioblastoma. *PLoS One*. 2010; 5: e10210. doi: [10.1371/journal.pone.0010210](https://doi.org/10.1371/journal.pone.0010210) PMID: [20419098](https://pubmed.ncbi.nlm.nih.gov/20419098/)
33. de Bont JM, Kros JM, Passier MMCJ, Reddingius RE, Sillevius Smitt PAE, Luijckx TM, et al. Differential expression and prognostic significance of SOX genes in pediatric medulloblastoma and ependymoma identified by microarray analysis. *Neuro Oncol*. 2008; 10: 648–60. doi: [10.1215/15228517-2008-032](https://doi.org/10.1215/15228517-2008-032) PMID: [18577562](https://pubmed.ncbi.nlm.nih.gov/18577562/)
34. Zhang J, Liang Q, Lei Y, Yao M, Li L, Gao X, et al. SOX4 induces epithelial-mesenchymal transition and contributes to breast cancer progression. *Cancer Res*. 2012; 72: 4597–608. doi: [10.1158/0008-5472.CAN-12-1045](https://doi.org/10.1158/0008-5472.CAN-12-1045) PMID: [22787120](https://pubmed.ncbi.nlm.nih.gov/22787120/)
35. Dyrskjøt L, Ostenfeld MS, Bramsen JB, Silahatoglu AN, Lamy P, Ramanathan R, et al. Genomic profiling of microRNAs in bladder cancer: miR-129 is associated with poor outcome and promotes cell death in vitro. *Cancer Res*. 2009; 69: 4851–60. doi: [10.1158/0008-5472.CAN-08-4043](https://doi.org/10.1158/0008-5472.CAN-08-4043) PMID: [19487295](https://pubmed.ncbi.nlm.nih.gov/19487295/)
36. Yeh Y-M, Chuang C-M, Chao K-C, Wang L-H. MicroRNA-138 suppresses ovarian cancer cell invasion and metastasis by targeting SOX4 and HIF-1 $\alpha$ . *Int J Cancer*. 2013; 133: 867–78. doi: [10.1002/ijc.28086](https://doi.org/10.1002/ijc.28086) PMID: [23389731](https://pubmed.ncbi.nlm.nih.gov/23389731/)
37. Andersen CL, Christensen LL, Thorsen K, Schepeler T, Sørensen FB, Verspaget HW, et al. Dysregulation of the transcription factors SOX4, CBF $\beta$  and SMARCC1 correlates with outcome of colorectal cancer. *Br J Cancer*. *Cancer Research UK*; 2009; 100: 511–23. doi: [10.1038/sj.bjc.6604884](https://doi.org/10.1038/sj.bjc.6604884)
38. Ramezani-Rad P, Geng H, Hurtz C, Chan LN, Chen Z, Jumaa H, et al. SOX4 enables oncogenic survival signals in acute lymphoblastic leukemia. *Blood*. *American Society of Hematology*; 2013; 121: 148–55. doi: [10.1182/blood-2012-05-428938](https://doi.org/10.1182/blood-2012-05-428938)
39. Classon M, Harlow E. The retinoblastoma tumour suppressor in development and cancer. *Nat Rev Cancer*. 2002; 2: 910–7. doi: [10.1038/nrc950](https://doi.org/10.1038/nrc950) PMID: [12459729](https://pubmed.ncbi.nlm.nih.gov/12459729/)
40. Fujitomo T, Daigo Y, Matsuda K, Ueda K, Nakamura Y. Critical function for nuclear envelope protein TMEM209 in human pulmonary carcinogenesis. *Cancer Res*. 2012; 72: 4110–8. doi: [10.1158/0008-5472.CAN-12-0159](https://doi.org/10.1158/0008-5472.CAN-12-0159) PMID: [22719065](https://pubmed.ncbi.nlm.nih.gov/22719065/)
41. Paderova J, Orlic-Milacic M, Yoshimoto M, da Cunha Santos G, Gallie B, Squire JA. Novel 6p rearrangements and recurrent translocation breakpoints in retinoblastoma cell lines identified by spectral

- karyotyping and mBAND analyses. *Cancer Genet Cytogenet.* 2007; 179: 102–11. doi: [10.1016/j.cancergencyto.2007.08.014](https://doi.org/10.1016/j.cancergencyto.2007.08.014) PMID: [18036396](https://pubmed.ncbi.nlm.nih.gov/18036396/)
42. Xu XL, Lee TC, Ofori N, Cheng C, Liu A, Fang Y, et al. Tumor-associated retinal astrocytes promote retinoblastoma cell proliferation through production of IGFBP-5. *Am J Pathol.* 2010; 177: 424–35. doi: [10.2353/ajpath.2010.090512](https://doi.org/10.2353/ajpath.2010.090512) PMID: [20508032](https://pubmed.ncbi.nlm.nih.gov/20508032/)
  43. Staaf J, Vallon-Christersson J, Lindgren D, Juliusson G, Rosenquist R, Höglund M, et al. Normalization of Illumina Infinium whole-genome SNP data improves copy number estimates and allelic intensity ratios. *BMC Bioinformatics.* 2008; 9: 409. doi: [10.1186/1471-2105-9-409](https://doi.org/10.1186/1471-2105-9-409) PMID: [18831757](https://pubmed.ncbi.nlm.nih.gov/18831757/)
  44. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics.* 2004; 5: 557–72. doi: [10.1093/biostatistics/kxh008](https://doi.org/10.1093/biostatistics/kxh008) PMID: [15475419](https://pubmed.ncbi.nlm.nih.gov/15475419/)
  45. Staaf J, Lindgren D, Vallon-Christersson J, Isaksson A, Göransson H, Juliusson G, et al. Segmentation-based detection of allelic imbalance and loss-of-heterozygosity in cancer cells using whole genome SNP arrays. *Genome Biol.* 2008; 9: R136. doi: [10.1186/gb-2008-9-9-r136](https://doi.org/10.1186/gb-2008-9-9-r136) PMID: [18796136](https://pubmed.ncbi.nlm.nih.gov/18796136/)
  46. Karolchik D, Hinrichs AS, Kent WJ. The UCSC Genome Browser. *Curr Protoc Bioinformatics.* 2012; Chapter 1: Unit1.4. doi: [10.1002/0471250953.bi0104s40](https://doi.org/10.1002/0471250953.bi0104s40)
  47. Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics.* 2012; 28: 1919–20. doi: [10.1093/bioinformatics/bts277](https://doi.org/10.1093/bioinformatics/bts277) PMID: [22576172](https://pubmed.ncbi.nlm.nih.gov/22576172/)