G3 Genes | Genomes | Genetics

INVESTIGATION

# *Drosophila* Larval Brain Neoplasms Present Tumour-Type Dependent Genome Instability

**Fabrizio Rossi,* Camille Stephan-Otto Attolini,* Jose Luis Mosquera,* and Cayetano Gonzalez*,†,1**

*Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac, 10, 08028 Barcelona, Spain and †Catalan Institution for Research and Advanced Studies (ICREA), 08010 Barcelona, Spain

ORCID IDs: 0000-0001-7020-2175 (F.R.); 0000-0001-8045-320X (C.S.-O.A.); 0000-0002-5259-3608 (C.G.)

**ABSTRACT** Single nucleotide polymorphisms (SNPs) and copy number variants (CNVs) are found at different rates in human cancer. To determine if these genetic lesions appear in Drosophila tumors we have sequenced the genomes of 17 malignant neoplasms caused by mutations *in l(3)mbt*, *brat*, *aurA*, or *lgl*. We have found CNVs and SNPs in all the tumors. Tumor-linked CNVs range between 11 and 80 per sample, affecting between 92 and 1546 coding sequences. CNVs are in average less frequent in *l(3)mbt* than in *brat* lines. Nearly half of the CNVs fall within the 10 to 100Kb range, all tumor samples contain CNVs larger that 100 Kb and some have CNVs larger than 1Mb. The rates of tumor-linked SNPs change more than 20-fold depending on the tumor type: at late time points *brat*, *l(3)mbt*, and *aurA* and *lgl* lines present median values of SNPs/Mb of exome of 0.16, 0.48, and 3.6, respectively. Higher SNP rates are mostly accounted for by C > A transversions, which likely reflect enhanced oxidative stress conditions in the affected tumors. Both CNVs and SNPs turn over rapidly. We found no evidence for selection of a gene signature affected by CNVs or SNPs in the cohort. Altogether, our results show that the rates of CNVs and SNPs, as well as the distribution of CNV sizes in this cohort of Drosophila tumors are well within the range of those reported for human cancer. Genome instability is therefore inherent to Drosophila malignant neoplastic growth at a variable extent that is tumor type dependent.

**KEYWORDS**

*Drosophila* cancer model

copy-number variation

single nucleotide polymorphism

genome instability

somatic mutation

A wide range of tumor types can be experimentally induced in different organs in *Drosophila melanogaster* (Gonzalez 2013; Tipping and Perrimon 2014; Markstein *et al.* 2014; Figueroa-Clarevega and Bilder 2015; Bangi *et al.* 2016; Sonoshita and Cagan 2017). Many of these tumors are hyperplasias that present during larval development and eventually differentiate, but others behave as frankly malignant neoplasms that are refractory to differentiation signals, lethal to the host and immortal. The latter can be maintained through successive rounds of allograft in adult flies (Rossi and Gonzalez 2015).



doi: https://doi.org/10.1534/g3.117.300489

Manuscript received November 30, 2017; accepted for publication February 1, 2018; published Early Online February 26, 2018.



Supplemental Material is available online at http://www.g3journal.org/lookup/suppl/doi:10.1534/g3.117.300489/-/DC1

[1]Corresponding author: IRB-Barcelona, Parc Cientific de Barcelona, Cell Division Laboratory, C/ Baldiri Reixac 10-12, 08028 Barcelona, Spain. E-mail: gonzalez@irbbarcelona.org

In humans, the study of mutational landscapes in thousands of tumors has generated a large catalog of genomic lesions that appear during tumor development and are a driving force for malignant growth in different cancer types (Redon *et al.* 2006; Stratton 2011; Vogelstein *et al.* 2013; Zarrei *et al.* 2015; Sudmant *et al.* 2015; Auton *et al.* 2015). In Drosophila, the sequencing of a single tumor caused by the loss of Polyhomeotic (Ph) revealed that neither single nucleotide polymorphisms (SNPs) nor copy number variations (CNVs) were significantly increased in comparison with non-tumoural control tissue, suggesting that genome instability (GI) may not be a pre-requisite for neoplastic epithelial growth in this model system. (Sievers *et al.* 2014). The question remains, however, as to the extent of GI in other samples of Ph tumors and, indeed, in different types of Drosophila malignant neoplasms.

To address this question we have investigated the mutational landscape of a cohort of tumors caused by mutations in *l(3)malignant brain tumor* (*l(3)mbt*), *brain tumor* (*brat*), *aurora-A* (*aurA*), and *l(2)-giant larvae* (*lgl*), which are some of the most aggressive and best characterized larval brain tumors that can be induced in Drosophila (Wright *et al.* 1976; Gateff 1978; Humbert *et al.* 2003; Lee *et al.* 2006; Wang *et al.* 2006; Bowman *et al.* 2008). Although similar in appearance under the dissection microscope, these tumors develop through different oncogenic pathways and originate from different cell types. Mutants

in *brat*, *aurA*, and *lgl* disrupt different aspects of the mechanisms of neuroblasts asymmetric division. The cell-of-origin of tumors caused by mutation in *brat* tumors is only the type II neuroblast, which resides in the dorsal side of the central brain (Bowman *et al.* 2008), while *aurA* and *lgl* tumors originate from type I and II neuroblasts (Gateff 1978; Humbert *et al.* 2003; Lee *et al.* 2006; Wang *et al.* 2006). Neoplastic growth in *l(3)mbt* tumors originate in the neuroepithelial regions of the larval brain lobes (Gateff 1978; Richter *et al.* 2011) and is tightly linked to the ectopic expression in the soma of germline genes (Janic *et al.* 2010).

Altogether, we sequenced a total of 17 genomes corresponding to a combination of tumor types, lines of the same tumor type, lines from the same individual, and time points. Our results show that CNVs and SNPs appear in Drosophila malignant neoplasms at a rate that is tumor-type dependent and within the range reported for human cancer.

## MATERIALS AND METHODS

### Fly strains

All fly stocks and crosses were maintained in standard food medium at 25° unless otherwise specified. Flies carrying the following mutants and transgenes were used: *pUbiGFP-tub84B* and *pUbi-His2Av::EYFP* (Rebollo *et al.* 2004) $l(3)mbt^{ts1}$ (Yohn *et al.* 2003), $brat^{k06028}$ (Spradling *et al.* 1999), $aurA^{8839}$ (Lee *et al.* 2006), $l(2)gl^{4}$ (Gateff 1978). The genotypes of each of the tumor lines are as follows. Lines mbtL1 and mbtL2: *Df(1)y-ac* $w^{1118}$, *pUbi-His2Av::EYFP*, *pUbq-alpha-tub-84::GFP*; $l(3)mbt^{ts1}$. Lines bratL1 and bratL2*:* $P\{w^{+}$, *lacW}*$brat^{k06028}$ (on a $w^{+}$ background). Line *lgl*: $l(2)gl^{4}$ Line *aurA*: $w^{1118}$, *pUbi-His2Av::EYFP*, *pUbq-alpha-tub-84::GFP*; $aurA^{8839}$. To generate *l(3)mbt* tumor larvae were raised at 29°.

### Allografts and DNA isolations

Allografts were performed as previously described (Rossi and Gonzalez 2015) with minor modifications. Single optic lobes from 3rd instar larvae were dissected and injected into the abdomen of $w^{1118}$ adult females. Flies were monitored daily and tumors were dissected out when they filled the abdomen of the host. Dissected tumors were resuspended in 100µl of PBS. An aliquot of 5µl of the tumor cell suspension was re-implanted in a new host and the remaining 95µl were processed for DNA isolation by standard lysis-ethanol precipitation, RNAse treatment, and beads-purification (Agencourt AMPure XP, Beckman Coulter). DNA from non-tumor larval tissues was isolated following the same protocol.

### Pair-end DNA sequencing

DNA samples from tumors and their relative controls were processed in parallel. Genomic DNA of each sample was extracted and then fragmented randomly by sonication. After electrophoresis, DNA fragments of about 150-300 bp were purified. Adapter ligation and DNA cluster preparation are performed by Illumina Nextera DNA Sample Preparation KIT (Illumina), and tumor and non-tumoural controls were sequenced in parallel by Illumina Hiseq2000. We performed read quality control using the FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). All samples passed minimum quality requirements.

### Alignment and coverage computation and correction (for CNV analyses)

100 bp paired end reads were aligned to the dm6 Drosophila genome version using the STAR aligner (Dobin *et al.* 2013) with default parameters. Each chromosome was binned into 1000bp segments for which mean coverage was computed using the IGVtools software (Robinson *et al.* 2011). We detected uneven coverage for regions with different GC content levels. In order to correct for this bias we fitted a generalized linear model using the Tweedie family with parameter 1.5 and log link function as implemented in the "gam" function from the R statistical language package "mgcv". Residuals were used for all subsequent calculations.

### Filtering, normalization, segmentation and CNV calling

We downloaded mappability information for the dm3 genome version from and converted coordinates to the dm6 version using the liftOver tool in (Kent *et al.* 2002). Mean GC content was computed for each 1kb bin from the dm6 genome version.

Bins with mappability values of 0 and GC content below the lower 0.08 quantile were removed from the analysis. Corrected and filtered coverage was quantile normalized for all samples using the function "normalize.quantiles" from the "preprocessCore" R package. Genome segmentation was performed according to (Olshen *et al.* 2004) using the "segment" function as implemented in the "DNAcopy" R package "CGHcall". Segmentation and all subsequent steps were performed for each tumor type independently. For each comparison of interest, the ratio was computed between the sample and its corresponding control. Ratios were further normalized using the "normalize" function from the same package. p-value cutoff was set to 0.01 and a minimum of 3 standard deviations between segments. Segment means were normalized using the "postsegnormalize" function. We used the "CGHcall" function from the "CGHcall" package to classify segments into double deletion, single deletion, diploid, single amplification and high amplification. Default parameters were used throughout the analysis.

Gene annotation was performed using the "biomaRt" R package (Durinck *et al.* 2009) version from May 2015.

In order to obtain CNVs outside of URs we removed bins overlapping with these regions and repeated the segmentation step and CNV calling. Repeat masker regions were downloaded from UCSC for *Drosophila melanogaster* dm6 version. Under-replicated regions were obtained from (Yarosh and Spradling 2014).

### Alignment and read processing for SNP calling

Reads were aligned to the dm6 version of the Drosophila genome using the BWA software version 0.7.6A (Li and Durbin 2009) with default parameters. The resulting output was converted to the bam format and sorted using samtools version 0.1.19 (Li *et al.* 2009). We then proceeded to process the data with the software package GATK version 2.5-2 (DePristo *et al.* 2011) according to their recommended best practices and with default parameters. We used a database of known SNPs downloaded from http://e68.ensembl.org/Drosophila_melanogaster corresponding to the dm3 genome version and converted to the dm6 version using the liftOver tool. Each sample was pre-processed according to the following steps: removal of duplicates using picard version 1.92; realignment of reads around indels using the GATK package with functions RealignerTargetCreator and IndelRealigner; base recalibration with the SNP database mentioned above and the function BaseRecalibrator from GATK.

### Somatic mutation calling

Preprocessed files were used as input for the muTect software version 1.1.4 (Cibulskis *et al.* 2013) with default parameters. Each sample was paired with its corresponding control. Resulting somatic SNPs were annotated using the software SNPeff version 3.0 (Cingolani *et al.* 2012).

### SNP clustering

We counted the number of SNPs in windows of 50KB around each SNP detected by our method. We then performed a binomial test assuming a

constant probability of finding a SNP in every position of the genome. The total effective size was computed as the number of positions with sufficient information in order to call a SNP. The Mutect algorithm internally defines these positions.

### Gene Ontology Enrichment

Gene ontology enrichment analysis was performed at the Gene Onto Consortium Website (http://geneontology.org) querying the set of 1791 genes that are amplified in at least 1 sample and never deleted and the set of 1101 genes that are deleted in at least 1 sample and never amplified in our cohort. We compared our gene set to the GO cellular component and biological function complete Data Sets and using the Bonferroni correction for multiple testing.

### Statistical analyses

Unless otherwise stated all statistical test in this study were calculated by the Mann-Whitney test.

### Data availability

Strains used in this study are available upon request. File S1 contains the following figures and tables. Figure S1 in File S1: Sequence coverage and first draft map of CNVs. Table S1 in File S1: Catalogue of CNVs found in the cohort. Table S2 in File S1: GO analyses of genes affected by CNVs. Table S3 in File S1: Catalogue of SNPs found in the cohort. Table S4 in File S1: SNP cluster analyses. Table S5 in File S1: SNP types found in the cohort. Table S6 in File S1: Percentage of SNPs passed on to later time points. Raw sequencing data are available at ENA with the accession number: PRJEB25187

## RESULTS/DISCUSSION

To determine the extent of genome instability (GI) in Drosophila malignant neoplasms we generated a cohort from six different larval brain tumors including two *l(3)mbt* (mbtL1 and mbtL2), two *brat* (bratL1 and bratL2), one *aurA*, and one *lgl* (Figure 1). Following allografts into adult hosts (Rossi and Gonzalez 2015), gDNA samples were taken at T0 (first round of allograft), T5, and in some cases T10. One of the *l(3)mbt* tumor lines was split at T9 into two sub-lines that were cultured separately up to T10.

### CNVs are frequent in the Drosophila brain tumor cohort

To identify copy number variants (CNVs) that appear during tumor growth we compared the gDNA coverage from each tumor sample of the cohort to that of the larva in which each of these tumors originated. Based on the detection of Y chromosome-specific sequences and the ratio of X chromosome/autosomes coverage we concluded that mbtL1, mbtL2, and *lgl* tumor lines originated from male larvae while aur, bratL1, and bratL2 originated from females (Figure S1A in File S1). We could not sex the tumors before allografting because testis do not develop in some of these mutant larvae. Most (88%) of the identified CNVs correspond to gains clustered on heterochromatic and under-replicated euchromatic regions (URs), which are present in all lines from T0. These regions do not endoreplicate to the full extent that most of the genomic DNA does in polytene larval tissues (Belyakin *et al.* 2005; Sher *et al.* 2012; Nordman and Orr-Weaver 2012; Yarosh and Spradling 2014) and therefore appear as copy number gains when the non-polytene tumor samples are compared to larval gDNA (Figure S1B in File S1). Their detection provides a valuable internal control for our CNV calling method. Running the algorithm after filtering these regions out with a repeat mask generates the map revealing the actual extent of CNVs that arise during tumor development in our cohort (Table S1 in File S1). A

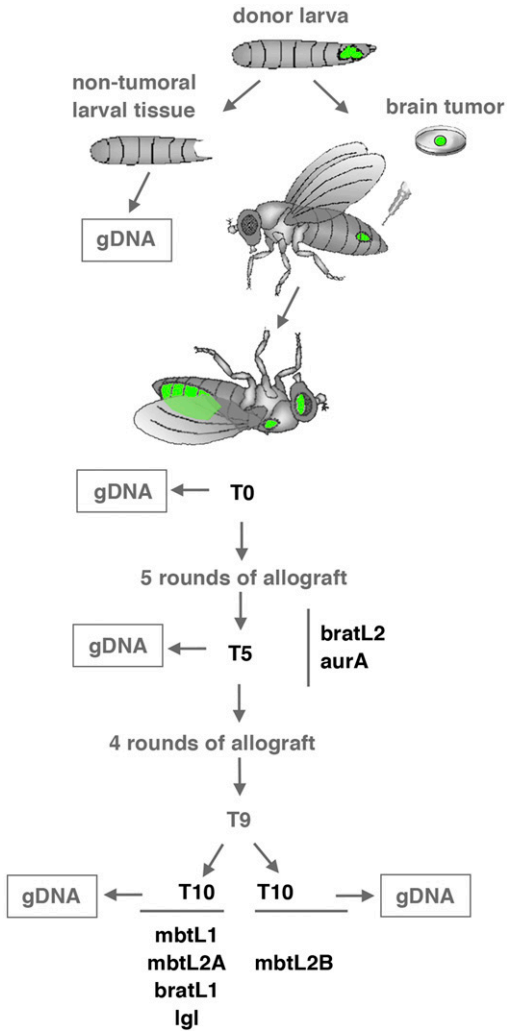


**Figure 1** The Drosophila larval brain tumor cohort. Larval brains tumors derived from two *l(3)mbt* (mbtL1 and mbtL2), two *brat* (bratL1 and bratL2) one *aurA*, and one *lgl* individuals were dissected out from the donor larvae and allografted repeatedly, up to T5 for bratL2 and *aurA*, and up to T10 for mbtL1, mbtL2, bratL1, and *lgl*. Line mbtL2 was split at T9 to generate sublines mbtL2A and mbtL2B. Genomic DNA was obtained from all tumor lines at T0, T5, and T10 if available, as well as from the non-tumoural tissues of the corresponding donor larvae.

graphic summary of the map of gains (≥ +2, blue; +1,green) and losses (-1,red; -2, magenta) on each chromosome arm is shown in Figure 2A. This final filtered map is not only a clean version of the unfiltered; it also includes new CNVs that can only be identified thanks to the finer calibration of diploidy achieved by the algorithm following the removal of URs and heterochromatin.

We detected CNVs in all tumor samples at rates that range between 11 in mbtL2 T10A to 80 in bratL1 T10 (Figure 2A, B) with an average of 37 ± 20.5. Differences among tumor types are not major, but CNVs per genome are in average significantly fewer in *l(3)mbt* (20.1 ± 9.6) than in *brat* (56.2 ± 17.3; $P = 0.005$) lines. The average number of each CNVs class (-2, -1, +1, and ≥+2) per genome in the entire cohort is 2.8 ± 4.8, 13.5 ± 10.5, 20.6 ± 14, and 0.2 ± 0.7, respectively. The only cases of ≥+2 were observed in the bratL1 line. Gains and losses of a single copy (Figure 2B, classes +1 and -1; green and red, respectively) account for 92% of the found CNVs, with class +1 being more frequent

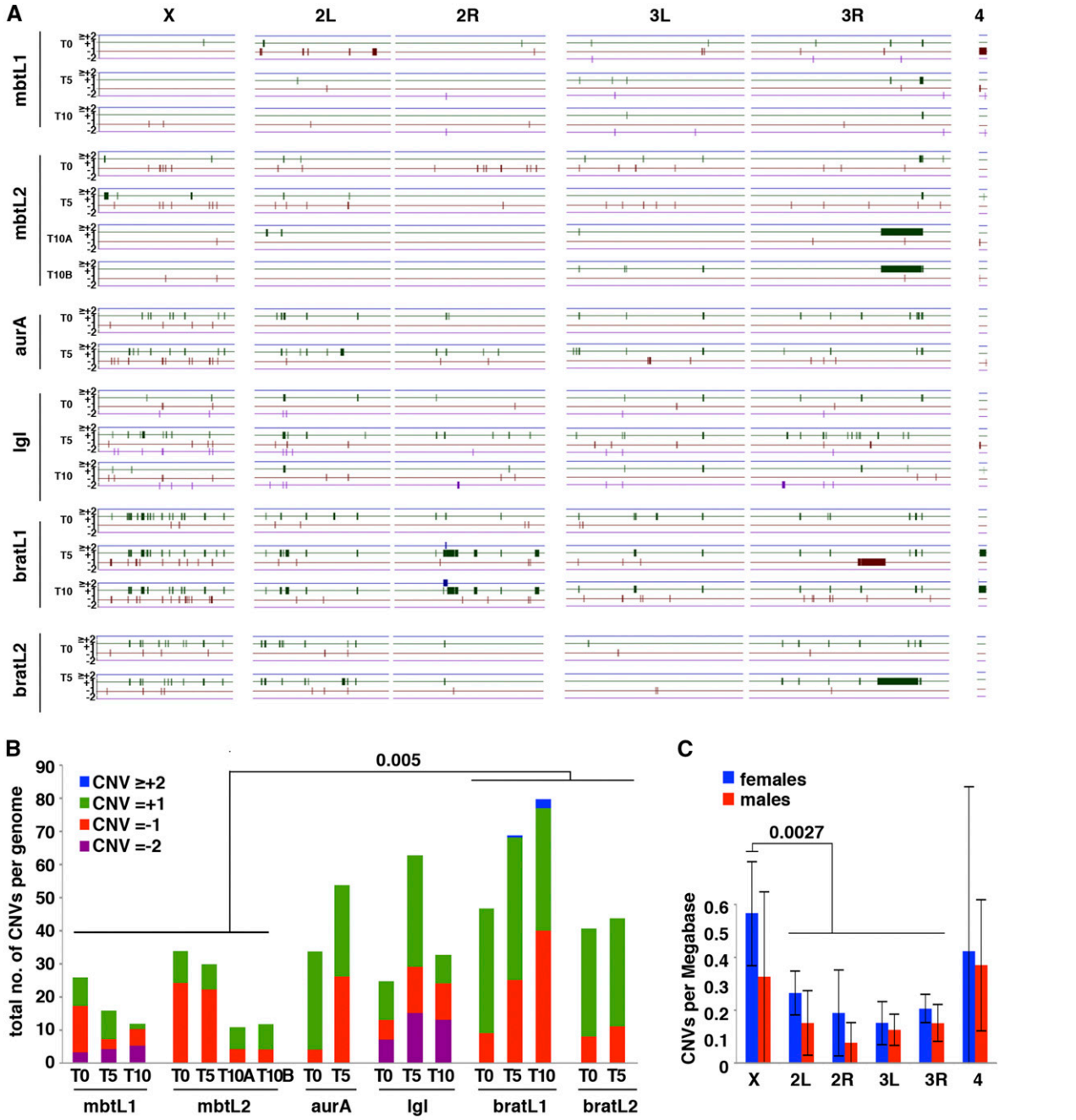


**Figure 2** Map and frequency of CNVs. A) Map of the CNVs identified in different lines at different time points after filtering out under-replicated regions. Gains (≥+2, blue; +1, green) and losses (-1, red; and ≤-2, purple) are mapped along chromosome arms X, 2L, 2R, 3L, 3R, and 4th. The heterochromatic Y chromosome is omitted. B) Barplot showing the total number of CNVs per genome per tumor sample and the relative contribution of each of the four CNV classes. C) Distribution of CNVs per Mb on each chromosome arm in female (blue) and male (red) samples. Error bars represent standard deviation.

in 65% or the samples. Amplifications are 1.3 times more abundant than deletions (354 and 277, respectively).

The four largest CNVs found in the cohort, much larger than all the rest, are one deletion and three duplications that, remarkably, fall in the same subdistal region in 3R and overlap extensively. The largest duplication was found in bratL2 T5 and spans 6.9 Mb on chromosome 3 (chr3R:20994001-27965000). This region (Figure 2A, longest thick green segment) overlaps extensively with two adjacent duplications of 4.0Mb (chr3R:21317001-2539800) and 2.5 Mb (chr3R:25402001-27960000) that are found in both mbtL2 T10A and mbtL2 T10B. Owed to the low resolution of Figure 2A, the two adjacent duplications appear as a single thick green segment in each tumor line. The large duplication in bratL2 T5 referred to above also overlaps over 1.1 Mb with the 4.1 Mb deletion (chr3R:17979001-22092000) observed in bratL1 T5, the largest deletion found in the cohort.

The rate of CNVs/Mb is slightly smaller in all chromosomes in male (range = 0.08-0.37 CNVs/Mb) than in female (range= 0.15-0.57) tumor samples, but differences are not significant (Figure 2C; $P = 0.055$). There are no significant differences in the rate of CNVs per Mb of

euchromatin among chromosomes, except for the X chromosome in female samples (0.57 ± 0.2 CNVs/Mb) which is significantly higher than in the autosomes ($P = 0.0027$).

All tumor samples in the cohort present a nearly diploid balance of chromosome stoichiometry (*i.e.*, 1X, 1Y, 2A in males; 2X, 2A in females). Most of the Y chromosome cannot be quantified due to the abundance of low complexity sequences and transposable elements (TEs). However, in all tumor samples derived from male larvae the coverage of the repeat-free *kl-2* gene region is very close to half of the mean coverage of the major autosomes, regardless of the time points of tumor growth. This result strongly suggests that unlike male cell lines, which often loose the entire Y (Lee *et al.* 2014), this chromosome is efficiently maintained in Drosophila tumors. The Y chromosome encodes only a handful of genes, all of them male fertility factors with no known function in the soma and, indeed, X/0 males are viable. However, the Y chromosome heterochromatin has a major impact on epigenetic variation and in modulating the expression of biologically relevant phenotypic variation (Lemos *et al.* 2010). Similarly, unlike Drosophila cell lines where widespread loss or gain of the entire chromosome 4 has been reported (Lee *et al.* 2014), we have only observed three cases of large segmental aneuplodies for this chromosome in our entire cohort: a deletion (-1) uncovering 93% of the euchromatin and two duplications (+1) covering 79% and 91% of the euchromatin of chromosome 4, respectively. As in many types of human cancer, karyotype changes have been observed in allografts from various larval brain tumors (Caussinus and Gonzalez 2005). In flies, karyotype changes do not appear to be sufficient on their own to drive tumourigenesis (Castellanos *et al.* 2008; Dekanty *et al.* 2012), but it is not known if they are involved in tumor progression. Our results suggest that specific aneuploid combinations are not selected during tumor progression.

## CNVs in tumor samples are larger than those found in Drosophila cell lines, and turn over rapidly

CNV size distribution is highly skewed and notably different between duplications and deletions (Figure 3A). Nearly half of the CNVs (49% of duplications and 47% of deficiencies) fall within the 10 to 100Kb range, but for those <10Kb, deletions and duplications account for 47% and 12% of the total, while in the >100Kb range the corresponding figures are 6% and 39% respectively. Indeed, most (85%, n = 20) of the largest CNVs (≥500Kb) are amplifications that appeared at or after T5 (Table S1 in File S1).

The total length of genomic sequences affected by gains in each tumor sample is quite significant, ranging between 180 Kb and 9.5 Mb. All but one of the 17 samples are affected by duplications covering more than 0.5 Mb. Deletions cover smaller, but still significant regions ranging from 60 Kb to 5.1 Mb. 15 out of 17 samples present deletions covering more than 100 Kb (Figure 3B). Genomic sequence length correlates tightly with the number of coding sequences affected by copy number variation (Figure 3B). In the entire cohort the number of genes affected by duplications or deletions range from 40 to 1404 and 9 to 773, respectively. In 11 out of the total 17 samples, duplications affect more than 100 genes and deletions affect more than 30 (Figure 3B).

Enrichment analysis of the genes duplicated in at least one sample and not deleted in any, shows only proteinaceous extracellular matrix (GO:0005578) as significantly overrepresented, and no GO term was found to be under-represented (Table S2 in File S1). Proteinaceous extracellular matrix is part of the GO term extracellular region (GO:0005576) that was found to be overrepresented in wild type strains (Dopman and Hartl 2007). Enrichment analysis of the genes deleted in at least one sample and not duplicated in any shows that the terms nucleosome assembly (GO:0006334), nuclear nucleosome (GO:0000788), and DNA-templated transcription initiation (GO:0006352), are significantly overrepresented, and no GO term was found to be under-represented (Table S2 in File S1). However, "nuclear function", which includes nuclear nucleosome and nucleosome assembly was found to be under-represented in duplicated fragments in wild type strains (Dopman and Hartl 2007).

The range of CNV sizes found in the tumor cohort is similar to those reported in Drosophila cell lines, and much larger than those found in wild type natural population and laboratory-adapted strains where 95% of the variants are shorter than 5 Kb and the largest duplicated and deleted regions are only 12 kb and 33 kb long, respectively (Dopman and Hartl 2007; Emerson *et al.* 2008; Cardoso-Moreira *et al.* 2012; Gilks *et al.* 2016). Moreover, unlike Drosophila strains where CNVs affect more frequently regions that do not contain coding sequences (Dopman and Hartl 2007; Emerson *et al.* 2008), 97% of the CNVs found in our tumor cohort affect coding sequences. The range of CNVs length in our tumor cohort is also much larger than those found in a Drosophila epithelial tumor caused by the loss of *polyhomeotic* (*ph*) (Sievers *et al.* 2014) and similar to the 0.5 kb – 85 Mb range found in human cancer (Beroukhim *et al.* 2010).

To get an estimate of the rate of turnover of CNVs, we plotted those that appear at any given T together with those that overlap in at least 1 Kb with CNVs found at the previous time point (Figure 3C). New variants, both amplifications and deletions, appear at each time point, but are diluted at a greater or lesser extent at later time points of tumor growth: the fraction of duplication and deletions passed on from T0 to T10 is within the 5–70% range, with no major differences between deficiencies and duplications. More than a third of the total number of CNVs found at any given T were not present at earlier time points. An interesting case reflecting the rate of CNV turnover is that of the pair mbtL2 T10A and T10B. These two samples, which were originated by splitting the mbtL2 line at T9, contained 11 and 12 CNVs respectively of which 8 were common to both lines, thus illustrating a case in which CNVs arise in a single round of transplantation. In total, deficiencies and duplications inherited from T0 account for 7 and 14% of those present at the last round of allograft, respectively.

Three main conclusions can be derived from our results. First, compared to those reported in Drosophila wild type strains, CNVs in our tumor cohort are much more abundant and larger and appear much faster, over a period of weeks rather than years. Such a high rate of interstitial aneuploidy strongly suggests that one or more of the pathways that prevent the formation of interstitial aneuploidies are significantly compromised in these tumors, more in *brat* than in *l(3)mbt*. Second, neither number nor size distribution appear to correlate with the time point of tumor growth. This observation strongly argues that the cause of the GI that originates CNVs is concomitant with the onset of neoplastic malignant growth. Finally, their rather random distribution among tumor types and rounds of allografting, rapid turn over, and absence of hotspots shared among different lines suggest that CNVs behave like passengers rather than drivers in these tumors.

## SNPs rates are tumor-type and tumor-age dependent

We used MuTect to call single nucleotide polymorphisms (SNPs) between each tumor sample and the non-tumoural tissues of the corresponding larvae (Figure 4A; Table S3 in File S1). SNPs in TEs or low complexity sequences were not taken into consideration for further analysis. We found SNPs in all tumor samples, at rates that are tumor type and tumor age-dependent. Total SNP numbers at T0

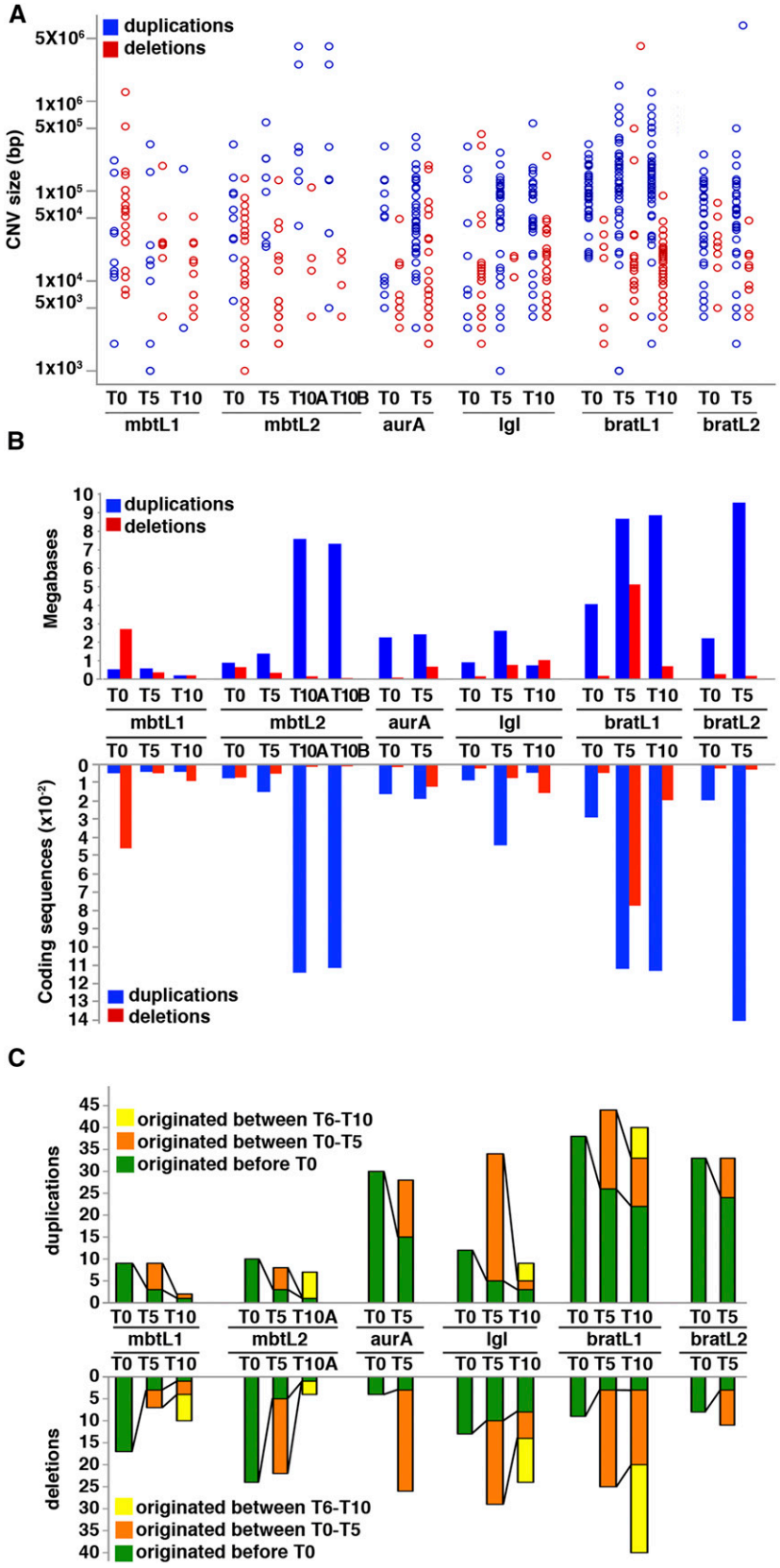


**Figure 3** Size distribution and turnover rate of CNVs. A, B) Distribution of CNVs sizes among the samples of the cohort. Duplications and deletions are shown in blue and red, respectively. A shows a scattered plot of CNV sizes in base-pairs, in logarithmic scale. B shows the total number of Mb (upper side of the graph) and total coding sequences (lower side of the graph) affected by duplications and deletions. C) Plot of number of duplications (upper side of the graph) and deletions (lower side of the graph) that are passed on through successive rounds of allograft.

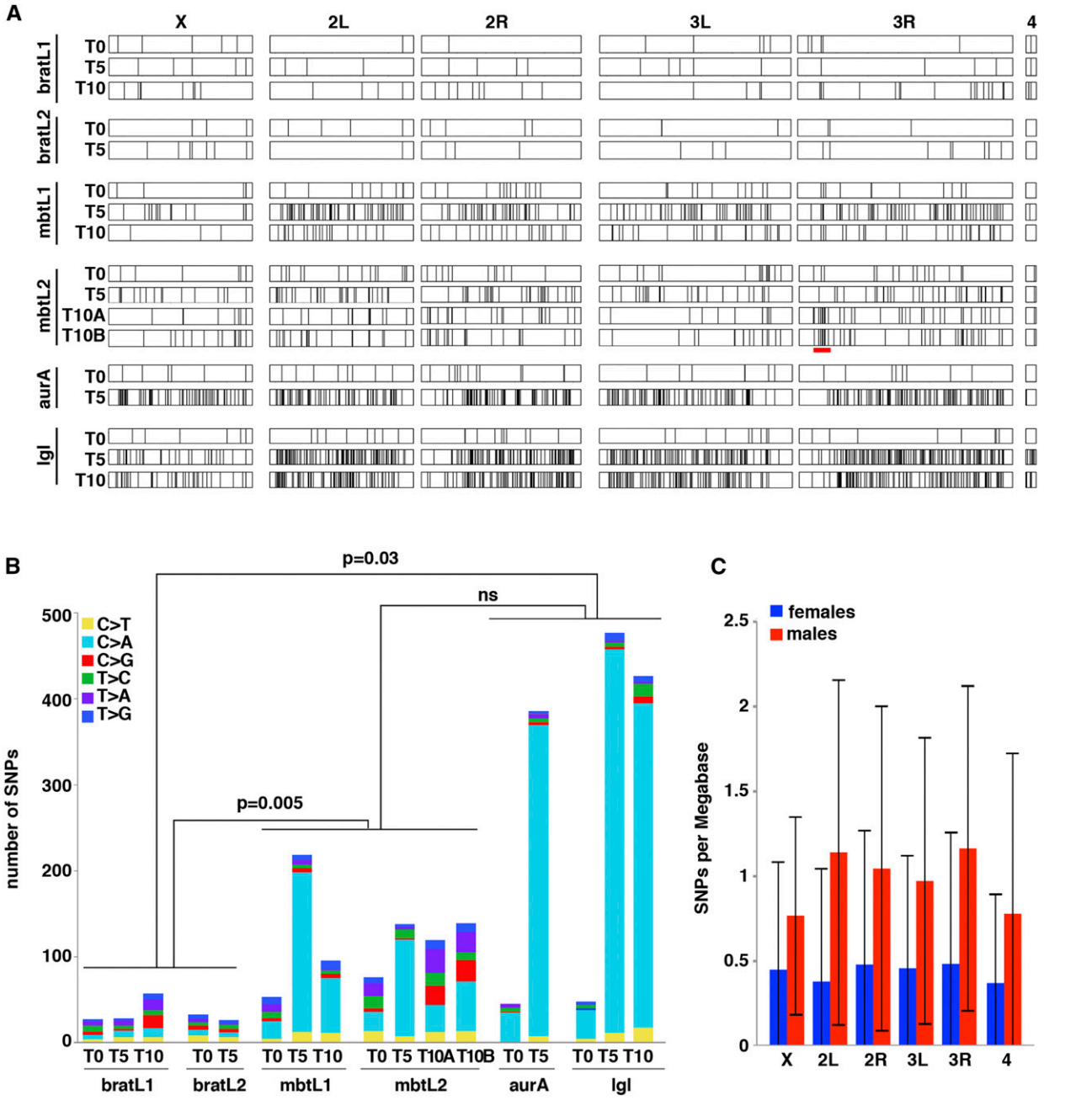


**Figure 4** Map and frequency of SNPs. A) The SNPs identified in different lines at different time points are mapped along chromosome arms X, 2L, 2R, 3L, 3R, and 4th. The heterochromatic Y chromosome is omitted. B) Barplot showing the total number of SNPs per genome per tumor sample and the relative contribution of each of the six possible base-pair substitutions. C) Distribution of SNPs per Mb on each chromosome arm in female (blue) and male (red) samples. Error bars represent standard deviation.

range between 27 and 76 among all tumor lines and remain unchanged at later time points in the two *brat* lines (range = 26-57). However, SNP burden increases to a range between 95 and 218 in the *l(3)mbt* lines and even more, up to eightfold compared to T0, in the *aurA* and *lgl* lines (range = 385-476)(Figure 4B). A previous report carried out by comparing tumor and control tissue to the Drosophila reference genome found no evidence of tumor-linked SNPs in one sample of allografted Ph tumor at T4 (Sievers *et al.* 2014). Using our own SNP calling strategy to directly compare the published tumor and control gDNA sequence we identify 20 tumor-linked SNPs, which is similar to the rate that we have found in the *brat* lines, the ones with the smallest number of SNPs within our cohort.

Most of the differences in the total number of SNPs among the tumor samples of our cohort are accounted for by C > A (G > T) transversions (Figure 4B, pale blue) to the extent that such differences among tumor lines at late time points become not significant if these two types of SNPs are removed. Indeed, the increase of C > A transversions becomes particularly notorious at later time points in *aurA* and *lgl* tumor lines where they account for more than 88% of all SNPs (Table S3 in File S1). Importantly, applying the method described by Costello *et al.* (Costello *et al.* 2013), we were able to discard the possible artifactual origin (*i.e.*, DNA oxidation during the processing of the DNA samples) of the C > A mutations that we have observed.

C > A (G > T) transversions are commonly produced by the formation of apurinic (abasic) sites or 8-hydroxy-2′-deoxyguanosine (8-oxo-dG) that result from superoxide anions reacting with deoxyguanosine (Mishra and Mishra 2002; De Bont and van Larebeke 2004). Because our sequencing data shows no evidence of mutations in genes involved in the removal of superoxide anions or 8-oxo-dG like *Sod2* (Kirby *et al.* 2002), *dOgg1* and *Ribosomal protein S3* (*RpS3*) (Dherin *et al.* 2000), or "DNA-(apurinic or apyrimidinic site) lyase activity", we hypothesize that the observed increase in C > A transversion may derive from tumor-type specific differences in metabolic activity and the consequent changes in oxidative stress levels.

The SNPs found in our cohort are scattered over the chromosomes and, unlike CNVs, they are not more frequent in the X chromosome than in the autosomes (Figure 4A, C). The lower rate of mean SNPs/Mb in all chromosomes in female samples may simply reflect the fact that the bratL1 and bratL2 lines, which present the lowest incidence of SNPs, are female and account for most (5/7) of the female samples of the cohort. By analyzing groups of SNPs separated by at most 50Kb we identified 96 regions where SNPs appear to be significantly ($P \leq 0.001$) clustered in each tumor line (Table S4 in File S1). However, none of our tumor samples showed any evidence of a "mutator phenotype" following a previously published method (Tamborero *et al.* 2013). The longest consecutive series of such clusters (about 400 Kb) maps to a chromosomal region that presents overall enrichment of SNPs, and that spans 3Mb in 3R. 17% (23/134) and 15% (22/151) of the SNPs found in the mbtL2 T10A and T10B lines, respectively, fall within this region, a highly significantly ($P \leq 1x10^{-12}$) increase compared to the 2% expected if SNPs were randomly distributed along the third chromosome.

## SNPs rates in Drosophila brain tumors are within the range reported for human tumors

To compare the rate of SNPs in our tumor cohort to those reported for human tumors (Lawrence *et al.* 2013) we determined the frequency of the various types of SNPs classified by their localization in the corresponding gene and deduced the rate of SNPs per Mb in the fraction of the exome that is sufficiently covered for significant SNP calling, considering only those SNPs with a minimum alternative allele frequency of 0.1 (Table S5 in File S1). For tumor lines with more than 100 SNPs, the fraction of SNPs falling in the exome ranges between 15 and 34% of which more than 60% affect protein sequence. The corresponding percentages are not significant in the lines that present fewer than 100 SNPs. Mean SNPs rate in *brat* tumors (0.16 SNP/Mb of exome) is close to that of the human tumors with the lowest rate of SNPs, like rhabdoid tumor (Figure 5). Mean SNPs rate in *l(3)mbt* tumors (0.48 SNP/Mb of exome) is within the range of pediatric medulloblastoma and neuroblastoma. Finally, the rates of SNPs in the exome in *aurA* and *lgl* (3.6 SNP/Mb of exome) fall among those of human tumors with a medium-high rate of SNPs, like glioblastoma multiforme, and colorectal cancers (Figure 5).

Malignancy traits are known to worsen over time in the tumors of our cohort: the later the round of implantation the higher the percentage of allografts that develop as tumors, and the shorter the life expectancy of implanted hosts (Caussinus and Gonzalez 2005; Castellanos *et al.* 2008). This observation strongly suggests the acquisition of driver mutations as tumors age. Such is the case in many human cancer types (Stratton *et al.* 2009; Vogelstein *et al.* 2013) as well as in established Drosophila cell lines which acquire pro-proliferation and anti-apoptotic mutations (Lee *et al.* 2014). However, we have found no genes mutated in more than one tumor line, not even among those with the highest rates of SNPs. Moreover, the fraction of SNPs that are passed on to later time points is very small ranging between 9 and 24 from T0 to T5 and between 0 and 8 from T5 to T10 (Table S6 in File S1). Thus, for instance, only 3% of the 476 SNPs found in lgl T5 were passed on to lgl T10. Altogether, these results do not support the presence of driver mutations in the cohort that we have analyzed. The point has to be made, however, that for detection of driver genes in human cancer, sample sizes are much larger than ours, in the order of hundreds per tumor type (Gonzalez-Perez *et al.* 2013). Therefore, the fact that our data does not reveal driver mutations in our cohort of Drosophila larval brain tumors does not rule out their existence.

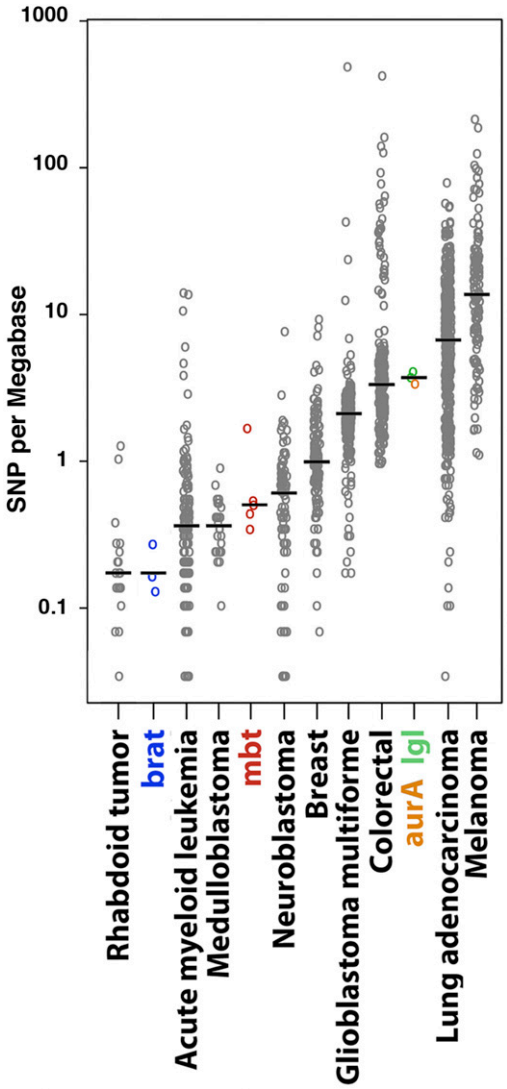


**Figure 5** SNP rates of Drosophila larval brain tumors compared to the SNP rate spectrum of a selection of human cancers. Scattered plot of the rates of SNPs/Mb of exome found in late time points (T5 and T10) of the Drosophila cohort (colored) together with those from a selection of human cancer samples (gray circles; modified after (Lawrence *et al.* 2013).

In summary, we have found that Drosophila larval brain malignant neoplasms with diverse origin present different SNP burdens that are well within the range of SNPs rates reported for human cancer. The very low percentage of SNPs passed on to later time points and the absence of genes mutated in more than one line strongly argues that, like CNVs, tumor-linked SNPs are passenger mutant. The very predominant

transversions are likely to result from enhanced oxidative stress conditions that are linked to tumor growth.


## ACKNOWLEDGMENTS

We thank the Bloomington Drosophila Stock Centre for providing fly stocks and A. Duran for technical assistance. We are also indebted to CNAG-CRG (Barcelona, Spain) for very helpful advice. Research in our laboratory is supported by ERC AdG 2011 294603 Advanced Grant from the European Research Council; BFU2015-66304-P and Redes de Excelencia BFU2014-52125-REDT-CellSYS from the Spanish MINECO, Spain; and SGR Agaur 2014 100 from Generalitat de Catalunya, Spain.


## LITERATURE CITED


Auton, A., L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang *et al.*, 2015 A global reference for human genetic variation. Nature 526(7571): 68–74. https://doi.org/10.1038/nature15393

Bangi, E., C. Murgia, A. G. Teague, O. J. Sansom, and R. L. Cagan, 2016 Functional exploration of colorectal cancer genomes using Drosophila. Nat. Commun. 7: 13615. https://doi.org/10.1038/ncomms13615

Belyakin, S. N., G. K. Christophides, A. A. Alekseyenko, E. V. Kriventseva, E. S. Belyaeva *et al.*, 2005 Genomic analysis of Drosophila chromosome underreplication reveals a link between replication control and transcriptional territories. Proc. Natl. Acad. Sci. USA 102(23): 8269–8274. https://doi.org/10.1073/pnas.0502702102

Beroukhim, R., C. H. Mermel, D. Porter, G. Wei, S. Raychaudhuri *et al.*, 2010 The landscape of somatic copy-number alteration across human cancers. Nature 463(7283): 899–905. https://doi.org/10.1038/nature08822

Bowman, S. K., V. Rolland, J. Betschinger, K. A. Kinsey, G. Emery *et al.*, 2008 The tumor suppressors Brat and Numb regulate transit-amplifying neuroblast lineages in Drosophila. Dev. Cell 14(4): 535–546. https://doi.org/10.1016/j.devcel.2008.03.004

Cardoso-Moreira, M., J. R. Arguello, and A. G. Clark, 2012 Mutation spectrum of Drosophila CNVs revealed by breakpoint sequencing. Genome Biol. 13(12): R119. https://doi.org/10.1186/gb-2012-13-12-r119

Castellanos, E., P. Dominguez, and C. Gonzalez, 2008 Centrosome dysfunction in Drosophila neural stem cells causes tumors that are not due to genome instability. Curr. Biol. 18(16): 1209–1214. https://doi.org/10.1016/j.cub.2008.07.029

Caussinus, E., and C. Gonzalez, 2005 Induction of tumor growth by altered stem-cell asymmetric division in Drosophila melanogaster. Nat. Genet. 37 (10): 1125–1129. https://doi.org/10.1038/ng1632

Cibulskis, K., M. S. Lawrence, S. L. Carter, A. Sivachenko, D. Jaffe *et al.*, 2013 Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. 31(3): 213–219. https://doi.org/10.1038/nbt.2514

Cingolani, P., A. Platts, L. Wang le, M. Coon, T. Nguyen *et al.*, 2012 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6(2): 80–92. https://doi.org/10.4161/fly.19695

Costello, M., T. J. Pugh, T. J. Fennell, C. Stewart, L. Lichtenstein *et al.*, 2013 Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic Acids Res. 41(6): e67. https://doi.org/10.1093/nar/gks1443

De Bont, R., and N. van Larebeke, 2004 Endogenous DNA damage in humans: a review of quantitative data. Mutagenesis 19(3): 169–185. https://doi.org/10.1093/mutage/geh025

Dekanty, A., L. Barrio, M. Muzzopappa, H. Auer, and M. Milan, 2012 Aneuploidy-induced delaminating cells drive tumorigenesis in Drosophila epithelia. Proc. Natl. Acad. Sci. USA 109(50): 20549–20554. https://doi.org/10.1073/pnas.1206675109

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire *et al.*, 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43(5): 491–498. https://doi.org/10.1038/ng.806

Dherin, C., M. Dizdaroglu, H. Doerflinger, S. Boiteux, and J. P. Radicella, 2000 Repair of oxidative DNA damage in Drosophila melanogaster: identification and characterization of dOgg1, a second DNA glycosylase activity for 8-hydroxyguanine and formamidopyrimidines. Nucleic Acids Res. 28(23): 4583–4592. https://doi.org/10.1093/nar/28.23.4583

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski *et al.*, 2013 STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29(1): 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dopman, E. B., and D. L. Hartl, 2007 A portrait of copy-number polymorphism in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA 104(50): 19920–19925. https://doi.org/10.1073/pnas.0709888104

Durinck, S., P. T. Spellman, E. Birney, and W. Huber, 2009 Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4(8): 1184–1191. https://doi.org/10.1038/nprot.2009.97

Emerson, J. J., M. Cardoso-Moreira, J. O. Borevitz, and M. Long, 2008 Natural selection shapes genome-wide patterns of copy-number polymorphism in Drosophila melanogaster. Science 320(5883): 1629–1631. https://doi.org/10.1126/science.1158078

Figueroa-Clarevega, A., and D. Bilder, 2015 Malignant Drosophila Tumors Interrupt Insulin Signaling to Induce Cachexia-like Wasting. Dev. Cell 33(1): 47–55. https://doi.org/10.1016/j.devcel.2015.03.001

Gateff, E., 1978 Malignant neoplasms of genetic origin in Drosophila melanogaster. Science 200(4349): 1448–1459. https://doi.org/10.1126/science.96525

Gilks, W. P., T. M. Pennell, I. Flis, M. T. Webster, and E. H. Morrow, 2016 Whole genome resequencing of a laboratory-adapted Drosophila melanogaster population sample. F1000 Res. 5: 2644.

Gonzalez, C., 2013 Drosophila melanogaster: a model and a tool to investigate malignancy and identify new therapeutics. Nat. Rev. Cancer 13(3): 172–183. https://doi.org/10.1038/nrc3461

Gonzalez-Perez, A., C. Perez-Llamas, J. Deu-Pons, D. Tamborero, M. P. Schroeder *et al.*, 2013 IntOGen-mutations identifies cancer drivers across tumor types. Nat. Methods 10(11): 1081–1082. https://doi.org/10.1038/nmeth.2642

Humbert, P., S. Russell, and H. Richardson, 2003 Dlg, Scribble and Lgl in cell polarity, cell proliferation and cancer. BioEssays 25(6): 542–553. https://doi.org/10.1002/bies.10286

Janic, A., L. Mendizabal, S. Llamazares, D. Rossell, and C. Gonzalez, 2010 Ectopic expression of germline genes drives malignant brain tumor growth in Drosophila. Science 330(6012): 1824–1827. https://doi.org/10.1126/science.1195481

Kent, W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle *et al.*, 2002 The human genome browser at UCSC. Genome Res. 12(6): 996–1006. https://doi.org/10.1101/gr.229102

Kirby, K., J. Hu, A. J. Hilliker, and J. P. Phillips, 2002 RNA interference-mediated silencing of Sod2 in Drosophila leads to early adult-onset mortality and elevated endogenous oxidative stress. Proc. Natl. Acad. Sci. USA 99(25): 16162–16167. https://doi.org/10.1073/pnas.252342899

Lawrence, M. S., P. Stojanov, P. Polak, G. V. Kryukov, K. Cibulskis *et al.*, 2013 Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 499(7457): 214–218. https://doi.org/10.1038/nature12213

Lee, C. Y., R. O. Andersen, C. Cabernard, L. Manning, K. D. Tran *et al.*, 2006 Drosophila Aurora-A kinase inhibits neuroblast self-renewal by regulating aPKC/Numb cortical polarity and spindle orientation. Genes Dev. 20(24): 3464–3474. https://doi.org/10.1101/gad.1489406

Lee, H., C. J. McManus, D. Y. Cho, M. Eaton, F. Renda *et al.*, 2014 DNA copy number evolution in Drosophila cell lines. Genome Biol. 15(8): R70. https://doi.org/10.1186/gb-2014-15-8-r70

Lemos, B., A. T. Branco, and D. L. Hartl, 2010 Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. Proc. Natl. Acad. Sci. USA 107(36): 15826–15831. https://doi.org/10.1073/pnas.1010383107

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14): 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16): 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Markstein, M., S. Dettorre, J. Cho, R. A. Neumuller, S. Craig-Muller *et al.*, 2014 Systematic screen of chemotherapeutics in Drosophila stem cell tumors. Proc. Natl. Acad. Sci. USA 111(12): 4530–4535. https://doi.org/10.1073/pnas.1401160111

Mishra, S. K., and P. C. Mishra, 2002 An ab initio theoretical study of electronic structure and properties of 2′-deoxyguanosine in gas phase and aqueous media. J. Comput. Chem. 23(5): 530–540. https://doi.org/10.1002/jcc.10046

Nordman, J., and T. L. Orr-Weaver, 2012 Regulation of DNA replication during development. Development 139(3): 455–464. https://doi.org/10.1242/dev.061838

Olshen, A. B., E. S. Venkatraman, R. Lucito, and M. Wigler, 2004 Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 5(4): 557–572. https://doi.org/10.1093/biostatistics/kxh008

Rebollo, E., S. Llamazares, J. Reina, and C. Gonzalez, 2004 Contribution of noncentrosomal microtubules to spindle assembly in Drosophila spermatocytes. PLoS Biol. 2(1): E8. https://doi.org/10.1371/journal.pbio.0020008

Redon, R., S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry *et al.*, 2006 Global variation in copy number in the human genome. Nature 444(7118): 444–454. https://doi.org/10.1038/nature05329

Richter, C., K. Oktaba, J. Steinmann, J. Muller, and J. A. Knoblich, 2011 The tumour suppressor L(3)mbt inhibits neuroepithelial proliferation and acts on insulator elements. Nat. Cell Biol. 13(9): 1029–1039. https://doi.org/10.1038/ncb2306

Robinson, J. T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E. S. Lander *et al.*, 2011 Integrative genomics viewer. Nat. Biotechnol. 29(1): 24–26. https://doi.org/10.1038/nbt.1754

Rossi, F., and C. Gonzalez, 2015 Studying tumor growth in Drosophila using the tissue allograft method. Nat. Protoc. 10(10): 1525–1534. https://doi.org/10.1038/nprot.2015.096

Sher, N., G. W. Bell, S. Li, J. Nordman, T. Eng *et al.*, 2012 Developmental control of gene copy number by repression of replication initiation and fork progression. Genome Res. 22(1): 64–75. https://doi.org/10.1101/gr.126003.111

Sievers, C., F. Comoglio, M. Seimiya, G. Merdes, and R. Paro, 2014 A deterministic analysis of genome integrity during neoplastic growth in Drosophila. PLoS One 9(2): e87090. https://doi.org/10.1371/journal.pone.0087090

Sonoshita, M., and R. L. Cagan, 2017 Modeling human cancers in Drosophila. Curr. Top. Dev. Biol. 121: 287–309. https://doi.org/10.1016/bs.ctdb.2016.07.008

Spradling, A. C., D. Stern, A. Beaton, E. J. Rhem, T. Laverty *et al.*, 1999 The Berkeley Drosophila Genome Project Gene Disruption Project: Single P-Element Insertions Mutating 25% of Vital Drosophila Genes. Genetics 153: 42.

Stratton, M. R., 2011 Exploring the genomes of cancer cells: progress and promise. Science 331(6024): 1553–1558. https://doi.org/10.1126/science.1204040

Stratton, M. R., P. J. Campbell, and P. A. Futreal, 2009 The cancer genome. Nature 458(7239): 719–724. https://doi.org/10.1038/nature07943

Sudmant, P. H., T. Rausch, E. J. Gardner, R. E. Handsaker, A. Abyzov *et al.*, 2015 An integrated map of structural variation in 2,504 human genomes. Nature 526(7571): 75–81. https://doi.org/10.1038/nature15394

Tamborero, D., A. Gonzalez-Perez, C. Perez-Llamas, J. Deu-Pons, C. Kandoth *et al.*, 2013 Comprehensive identification of mutational cancer driver genes across 12 tumor types. Sci. Rep. 3(1): 2650. https://doi.org/10.1038/srep02650

Tipping, M., and N. Perrimon, 2014 Drosophila as a model for context-dependent tumorigenesis. J. Cell. Physiol. 229(1): 27–33.

Vogelstein, B., N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, Jr *et al.*, 2013 Cancer genome landscapes. Science 339(6127): 1546–1558. https://doi.org/10.1126/science.1235122

Wang, H., G. W. Somers, A. Bashirullah, U. Heberlein, F. Yu *et al.*, 2006 Aurora-A acts as a tumor suppressor and regulates self-renewal of Drosophila neuroblasts. Genes Dev. 20(24): 3453–3463. https://doi.org/10.1101/gad.1487506

Wright, T. R., G. C. Bewley, and A. F. Sherald, 1976 The genetics of dopa decarboxylase in Drosophila melanogaster. II. Isolation and characterization of dopa-decarboxylase-deficient mutants and their relationship to the alpha-methyl-dopa-hypersensitive mutants. Genetics 84(2): 287–310.

Yarosh, W., and A. C. Spradling, 2014 Incomplete replication generates somatic DNA alterations within Drosophila polytene salivary gland cells. Genes Dev. 28(16): 1840–1855. https://doi.org/10.1101/gad.245811.114

Yohn, C. B., L. Pusateri, V. Barbosa, and R. Lehmann, 2003 l(3)malignant brain tumor and three novel genes are required for Drosophila germ-cell formation. Genetics 165: 1889–1900.

Zarrei, M., J. R. MacDonald, D. Merico, and S. W. Scherer, 2015 A copy number variation map of the human genome. Nat. Rev. Genet. 16(3): 172–183. https://doi.org/10.1038/nrg3871

*Communicating editor: M. Boutros*