GeTallele: a method for integrative analysis and visualization of DNA and RNA allele frequencies

Piotr Słowiński[1,#,*], Muzi Li[2,#], Paula Restrepo[2,3], Nawaf Alomran[2], Liam F. Spurr[2,4,5], Christian Miller[2], Krasimira Tsaneva-Atanasova[1,6] and Anelia Horvath[2,7,8]

[1]Department of Mathematics, College of Engineering, Mathematics and Physical Sciences, Living Systems Institute, Translational Research Exchange @ Exeter and EPSRC Centre for Predictive Modelling in Healthcare, University of Exeter, EX4 4QJ Exeter, UK, [2]McCormick Genomics and Proteomics Center, School of Medicine and Health Sciences, The George Washington University, 20037 Washington, DC, USA, [3]Department of Genetics & Genomics Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, [4]Cancer Program, Broad Institute of MIT and Harvard, Cambridge, MA, [5]Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, [6]Department of Bioinformatics and Mathematical Modelling, Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria, [7]Department of Pharmacology and Physiology, School of Medicine and Health Sciences, The George Washington University, 20037 Washington, DC, USA, [8]Department of Biochemistry and Molecular Medicine, School of Medicine and Health Sciences, The George Washington University, 20037 Washington, DC, USA.

* To whom correspondence should be addressed.

# Equal contribution

Abstract

Background: Asymmetric allele expression typically indicates functional and/or structural features associated with the underlying genetic variants. When integrated, RNA and DNA allele frequencies can reveal patterns characteristic of a wide-range of biological traits, including ploidy changes, genome admixture, allele-specific expression and gene-dosage transcriptional response.

Results: To assess RNA and DNA allele frequencies from matched sequencing datasets, we introduce GeTallele: a mathematical model and a toolbox that provides a suit of functions for integrative analysis, statistical assessment and visualization of **Ge**nome and **T**ranscriptome **allele** frequencies. The toolbox introduces a method for generating model distributions of variant allele frequencies (VAF) with a given variant read probability. Using model VAF probabilities, GeTallele allows estimation and comparison of variant read probabilities (VAF distributions) in a sequencing data sets. We demonstrate this functionality across cancer DNA and RNA sequencing data sets.

Conclusion: Based on our evaluation, variant read probabilities can serve as a dependable indicator to assess gene and chromosomal allele asymmetries and to aid calls of genomic events in matched sequencing RNA and DNA datasets.

Contact: P.M.Slowinski@exeter.ac.uk

1 Introduction

RNA and DNA carry and present genetic variation in related yet distinct manners; the differences being informative of functional and structural traits. In diploid organisms, an important measure of genetic variation is the allele frequency, which can be measured from both genome (DNA) and transcriptome (RNA) sequencing data (encoded and expressed allele frequency, respectively). Differential DNA-RNA allele frequencies are associated with a variety of biological processes, such as genome admixture or allele-specific transcriptional regulation (Ferreira, et al., 2016; Ha, et al., 2012; Han, et al., 2015; Movassagh, et al., 2016; Shah, et al., 2012).

Most of the RNA-DNA allele comparisons from sequencing have been approached at nucleotide level, where it has proven to be highly informative for determining the alleles' functionality (Ferreira, et al., 2016; Ha, et al., 2012; Han, et al., 2015; Macaulay, et al., 2016; Morin, et al., 2013; Movassagh, et al., 2016; Reuter, et al., 2016; Shah, et al., 2012; Shi, et al., 2016; Shlien, et al., 2016; The, et al., 2012; Yang, et al., 2016). Comparatively, integration of allele signals at the molecular level, as derived from linear DNA and RNA carriers, is less explored due to challenges presented by short sequencing length. The different molecular nature of RNA and DNA also leads to limited compatibility of the sequencing output.

Herein, we address some of the above challenges by employing a mathematical model to assess differences between RNA- and DNA- variant allele frequencies (VAF) at single nucleotide variant (SNV) positions in the genome. To do that, we introduce GeTallele: a toolbox that provides a suit of functions for integrative analysis and visualization of **Ge**nome (VAF_{DNA}) and **T**ranscriptome (VAF_{RNA}) **allele** frequencies along genes and chromosomes. Using VAF, GeTallele infers possible copy number alterations (CNAs), and, at

positions of interest, mathematically and statistically compares VAF_{RNA} and VAF_{DNA} distributions. Furthermore, GeTallele supports visualization of the allele distribution at a desired resolution—from nucleotide to genome. In contrast to other CNA-modelling methods based on statistical models, GeTallele is based on a mechanistic model of VAF distributions.

2 Results

We demonstrate GeTallele's functionality using sequencing DNA and RNA datasets from paired normal and tumour tissue obtained from 72 female patients with breast invasive carcinoma (BRCA) from The Cancer Genome Atlas (TCGA). Each dataset contains four matched sequencing sets: normal exome (Nex), normal transcriptome (Ntr), tumour exome (Tex), and tumour transcriptome (Ttr). The raw sequencing data were processed as previously described (Movassagh, et al., 2016) to generate the inputs for GetAllele.

The overall workflow of GetAllele is shown on Figure 1. As an input, GeTallele requires the absolute number of sequencing reads bearing the variant and reference nucleotide in each single-nucleotide variant (SNV) position across the four matched datasets. For each sample, to select SNV positions for analysis, we start with the high quality heterozygous SNV calls on the normal exome (Li, et al., 2009). In each of these positions we estimate the counts of the variant and reference reads (n_{VAR} and n_{REF}, respectively) across the 4 matching datasets, and retain for further analyses only positions covered by a minimum total (variant and reference) required unique sequencing reads across. This threshold is flexible and is required to ensure that only sufficiently covered positions will be analysed; it is set to 3 in the herein presented results.

From each of the 4 matched datasets GetAllele estimates VAF based on n_{VAR} and n_{REF} covering the positions of interest: $VAF = n_{VAR}/(n_{VAR}+n_{REF})$. An example of genome-wide VAF values estimated from Tex, and their corresponding distribution of VAF values is shown in Figure 2.

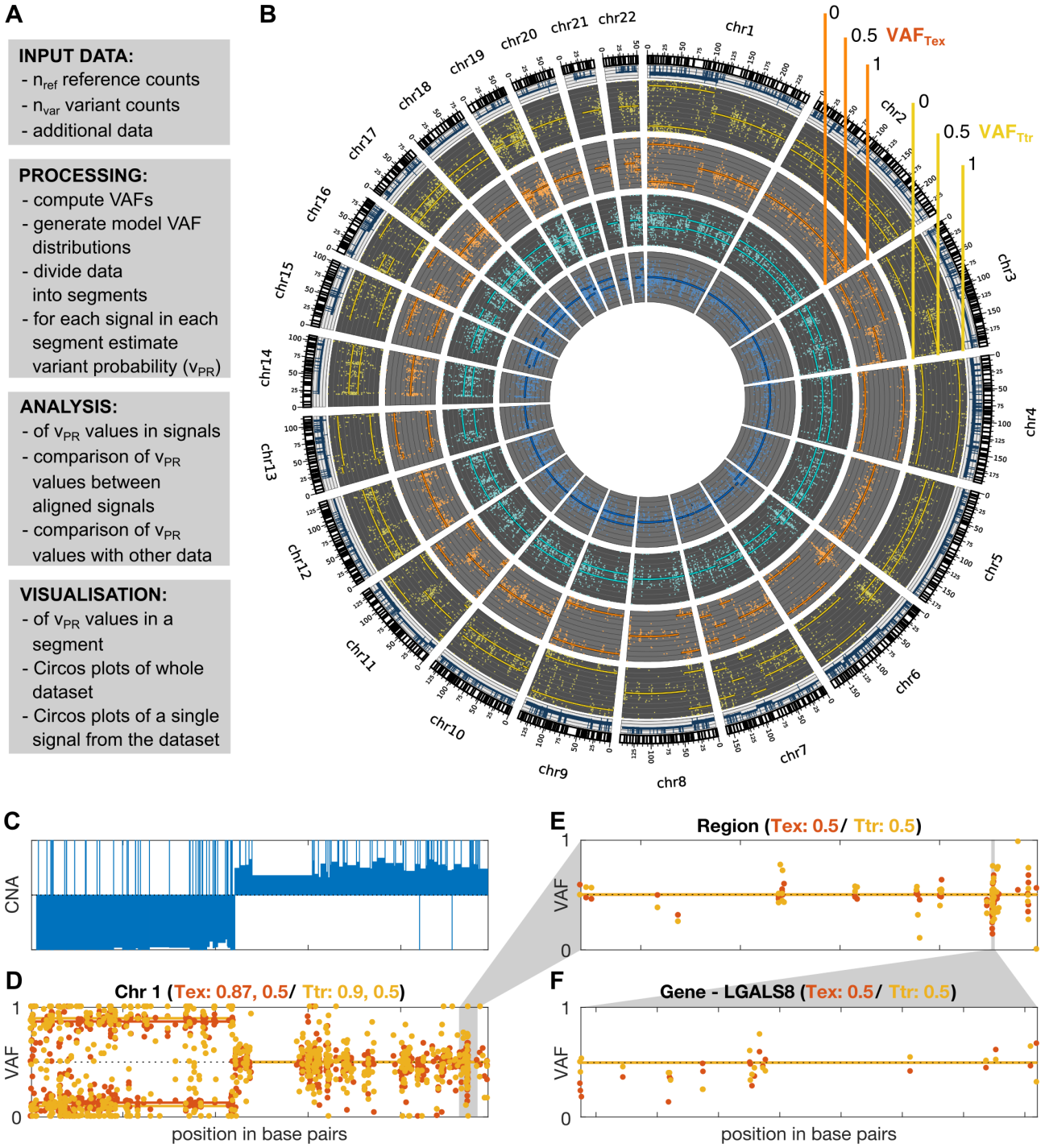


Fig. 1. GeTallele and visualisation of VAF data. A Toolbox description. B Visualisation of the whole dataset on the level of genome using Circos plot (blue – normal exome, cyan – normal transcriptome, orange - tumour exome, yellow - tumour transcriptome). C - F show in details VAF_{TEX} and VAF_{TTR} values of chromosome 1; C - F Visualization of the VAF values with fitted variant probability (v_{PR} – see Section 3.1 and Figure 2) values at the level of chromosome (D), custom genome region (E) and gene (F), for the chromosome level shown also are CNA values (C). Panel D shows that there are two segments with different VAF distributions. Panel C shows that change in the CNA is concurrent with the change in the VAF distributions. Tex - tumour exome (orange); Ttr - tumour transcriptome (yellow).

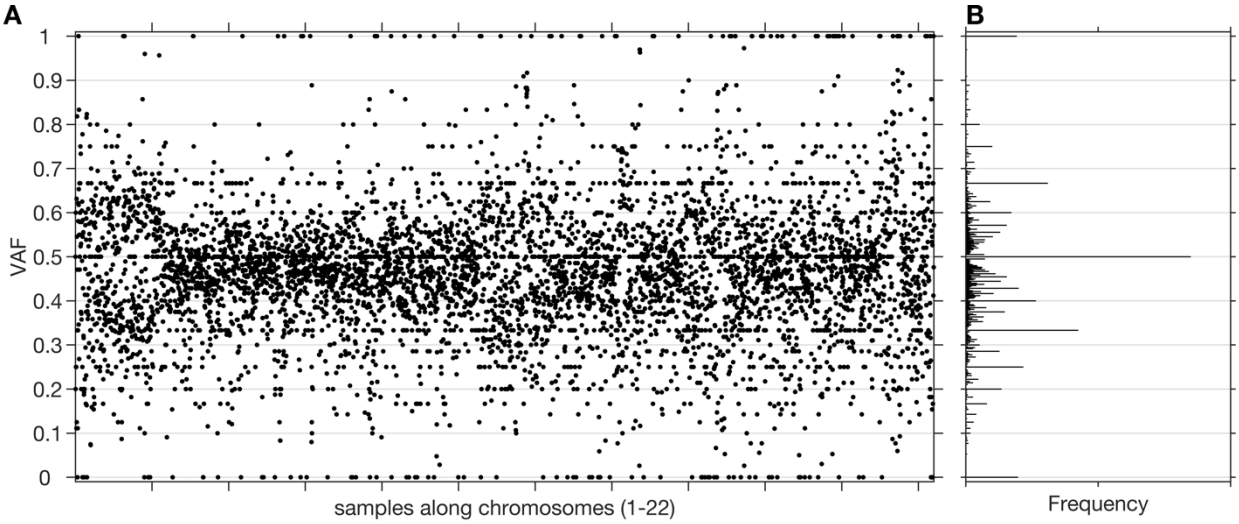


Fig. 2. Sample and distribution of variant allele frequencies (VAF) values. A. All the VAF values of a tumour exome sequencing signal (chromosomes 1-22) from one of the datasets. **B.** Histogram of the VAF values from panel A. Centres of bins of the histogram are located at elements of a Farey sequence.

Analysis of the VAF_{RNA} and VAF_{DNA} in the GeTallele is based on comparing probability of observing a given VAF value at various positions of interest. Estimation of the variant allele probability (v_{PR}) is implemented using a data-driven mathematical model of distribution of the VAF values and is the core functionality of the GeTallele.

2.1 Estimation of variant probability v_{pr}

2.1.1 Variant probability

GetAllele aggregates frequency counts of a VAF sample into variant probability, v_{PR} – probability of observing a variant allele. The v_{PR} is a single informative value indicative of copy number variation. The v_{PR} is a parameter that describes the genomic event that through the sequencing process was transformed into a specific distribution of VAF values found in the signal. For example, in VAF_{DNA} from a diploid genome, variant probability v_{PR}=0.5 (meaning that both alleles are equally probable) corresponds to a true allelic ratio of 1:1 for heterozygote sites. For heterozygote sites in the DNA from diploid monoclonal sample, the corresponding tumour VAF_{DNA} is expected to have the following interpretations: v_{PR}=1 or v_{PR}=0 corresponds to a monoallelic status resulting from a deletion, and v_{PR}=0.8 (or 0.2), 0.75 (or 0.25), 0.67 (or 0.33) correspond to allele-specific tetra-, tri-, and duplication of the variant-bearing allele, respectively.

The v_{PR} of the VAF_{RNA} is interpreted as follows. In positions corresponding to DNA heterozygote sites, alleles not preferentially targeted by regulatory traits are expected to have expression rates with variant probability v_{PR}=0.5, which (by default) scale with the DNA allele distribution. Differences between VAF_{DNA} and VAF_{RNA} values are observed in special cases of transcriptional regulation where one of the alleles is preferentially transcribed over the other. In the absence of allele-preferential transcription, VAF_{DNA} and VAF_{RNA} are anticipated to have similar v_{PR} across both diploid (normal) and aneuploid (affected by CNAs) genomic regions. Consequently, VAF_{DNA} and VAF_{RNA} are expected to synchronously switch between allelic patterns along the chromosomes, with the switches indicating break points of DNA deletions or amplifications.

Since we observed that DNA and RNA signals have different distributions of total reads and also that the distributions of total reads vary between participants, the model VAF distributions are generated individually for each sequencing signal and each participant.

To estimate v_{PR} in the signals, GeTallele first generates model VAF distributions and then uses the earth mover's distance (EMD) (Kantorovich and Rubinstein, 1958; Levina and Bickel, 2001) to fit them to the data. To generate a model VAF distribution with a given variant probability, v_{PR}, GeTallele, bootstraps 10000 values of the total reads (sum of the variant and reference reads; $n_{VAR} + n_{REF}$) from the analysed signal in the dataset. It then uses binomial pseudorandom number generator to get number of successes for given number of total reads and a given value of v_{PR} (implemented in the Matlab function `binornd`). The v_{PR} is the probability of success and generated number of successes is interpreted as an n_{VAR}. Since the model v_{PR} can take any value, it can correspond to a single genomic event as well as any combination of genomic events in any mixture of normal and tumour populations (See section 2.1.3).

The analysis presented in the paper uses 51 model VAF distributions with v_{PR} values that vary from 0.5 to 1 with step (increment of) 0.01. The model VAF distributions are parametrized using only $v_{PR} \geq 0.5$, however, to generate them we use v_{PR} and its symmetric counterpart $1-v_{PR}$. The process of generating model VAF distributions and examples of model and real VAF distributions with different values of v_{PR} are illustrated in Fig. 3.

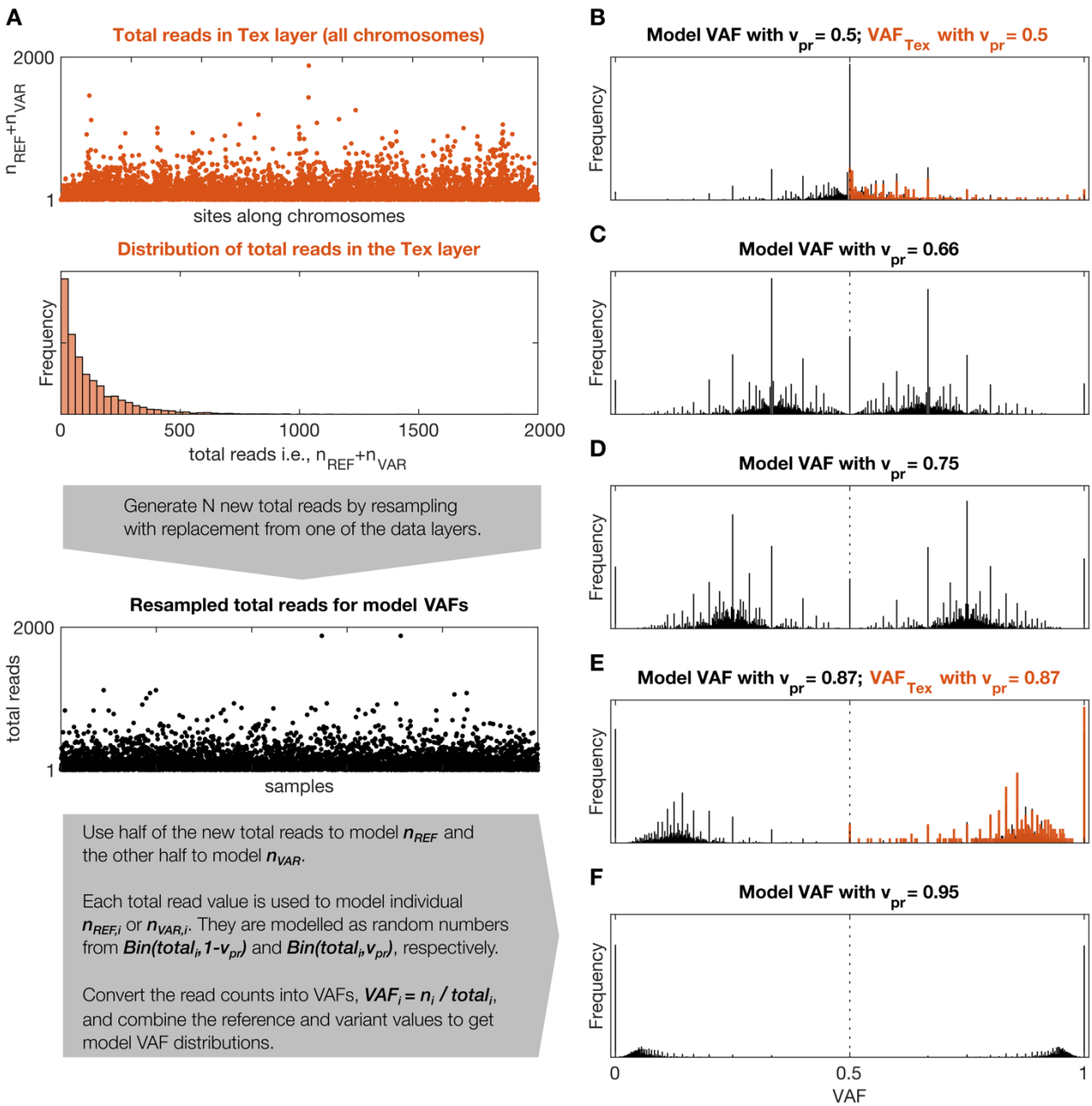


Fig. 3. Model and real VAF distributions. A Description of the model used to generate model VAF distributions, *Bin(n,p)* stands for binomial distribution with parameters *n* (number of trials) and *p* (probability of success). **B** - **F** Model VAF distributions for different values of v_{PR}. Panels **B** and **F** show additionally distributions of VAF_{TEX} for the two windows shown in Figure 1D.

2.1.2 v_{PR} estimation

Earth mover's distance (EMD) is a metric for quantifying differences between probability distributions (Kantorovich and Rubinstein, 1958; Levina and Bickel, 2001) and in the case of univariate distributions it can be computed as:

$$\mathrm{EMD}(\mathrm{PDF}_1,\ \mathrm{PDF}_2) = \int_Z |CDF_1(z) - CDF_2(z)| dz.$$

Here, PDF_1 and PDF_2 are two probability density functions and CDF_1 and CDF_2 are their respective cumulative distribution functions. Z is the support of the PDFs (i.e. set of all the possible values of the random variables described by them). Because VAFs are defined as simple fractions with values between 0 and 1, their support is given by a Farey sequence (Hardy, et al., 2008) of order n; n is the highest denominator in the sequence. For example, Farey sequence of order 2 is 0, 1/2, 1 and Farey sequence of order 3 is 0, 1/3, 1/2, 2/3, 1. GeTallele uses a Farey sequence of order 1000 as the support Z for all computations involving EMD.

To estimate v_{PR}, GeTallele computes EMD between the distribution of the VAF values of each signal in the window and the 51 model VAF distributions (i.e. observed vs modelled VAF), the estimate is given by the v_{PR} of the model VAF distribution that is closest to the VAF distribution in the window. Examples of VAF distributions with fitted model VAF distributions are shown in Figure 3A and D. The dependence of the confidence intervals of the estimation on the number of VAF values in a window is illustrated in Fig. 4, which clearly demonstrates that the larger the number of VAFs in the chosen window the higher the accuracy of the estimate.

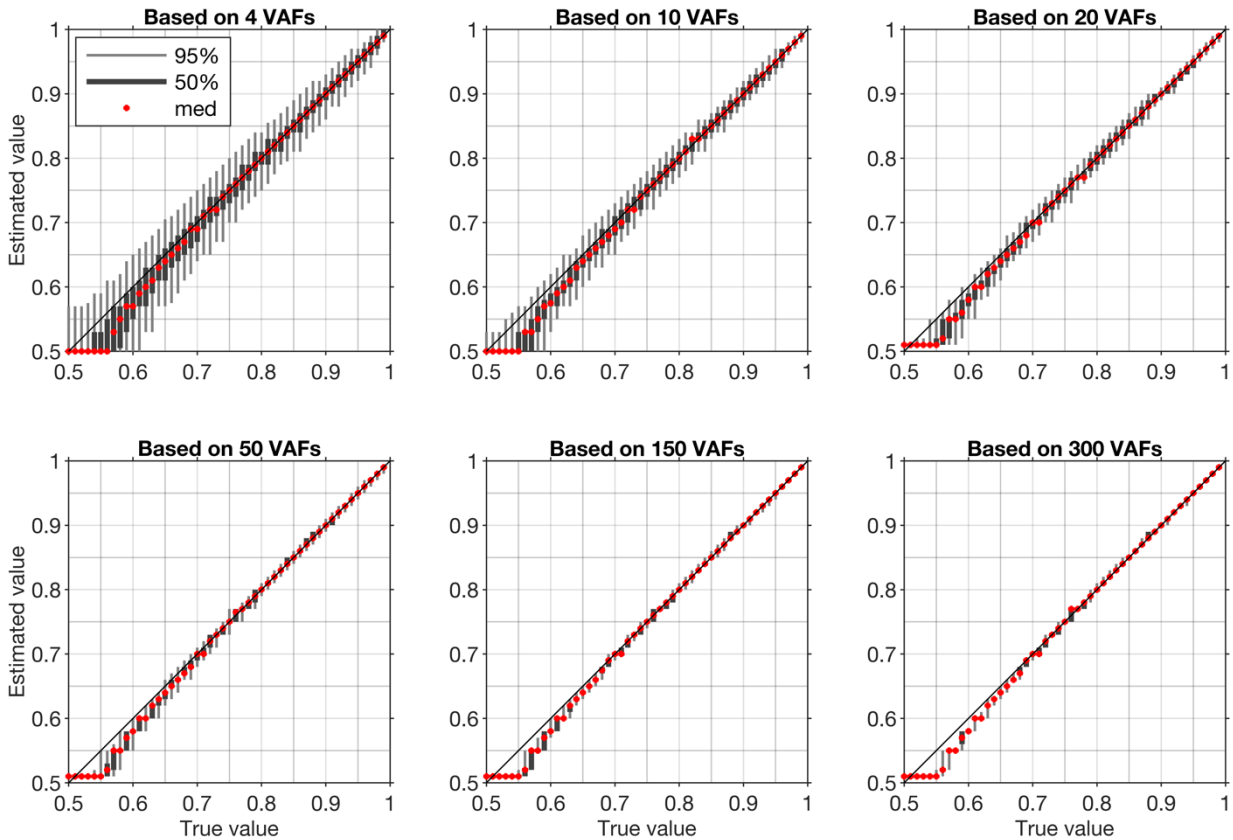


Fig. 4. Confidence intervals for artificial samples with different numbers of VAFs. Each confidence interval is based on estimation of v_{PR} in 1000 randomly generated samples with a fixed v_{PR} (True value). Light grey bar is 95% confidence interval (950 samples lay within this interval), dark grey bar is 50% confidence interval (500 samples lay within this interval), red dot is median value.

2.1.3 v_{PR} values in mixtures of normal and tumour populations

Since the v_{PR} can take any value between 0.5 and 1 it can correspond to a single genomic event as well as any combination of genomic events in any mixture of normal and tumour populations. A mixture v_{PR} value that corresponds to a combination of genomic events can be computed using the following expression:

$$v_{PR} = \frac{\sum_{pl=1}^{pl=N} \sum_{e_{VAR}=\{events\}} e_{VAR} \cdot p_{PL}}{\sum_{pl=1}^{pl=N} \sum_{e_{VAR}=\{events\}} e_{VAR} \cdot p_{PL} + \sum_{pl=1}^{pl=N} \sum_{e_{REF}=\{events\}} e_{REF} \cdot p_{PL}}$$

Where e_{VAR} and e_{REF} are the multiplicities of variant and reference alleles and p_{PL} is a proportion of one of the populations. For heterozygote sites e_{VAR}=1 and e_{REF}=1, for deletions e_{VAR}=0 or e_{REF}=0, for du-, tri- and

tetraplications e_{VAR} or e_{REF} can be equal to 2, 3 or 4, respectively. The sum of proportions p_{PL} over the populations is equal 1. For example, for a mixture of 1 normal (N, p_N=0.44) and 2 tumour populations (T1, p_{T1}=0.39 and T2, p_{T2}=0.17), T1 with deletion and T2 with deletion the mixture v_{PR} value can be computed as follows:

$$v_{PR} = \frac{p_N \cdot B + p_{T1} \cdot B + p_{T2} \cdot B}{p_N \cdot (A + B) + p_{T1} \cdot (0 + B) + p_{T2} \cdot (0 + B)}$$

$$= \frac{0.44 \cdot 1 + 0.39 \cdot 1 + 0.17 \cdot 1}{0.44 \cdot (1 + 1) + 0.39 \cdot (1 + 0) + 0.17 \cdot (1 + 0)} = 0.694.$$

By comparing the v_{PR} values estimated from data with possible mixture v_{PR} values we propose to estimate sample purity and its clonal composition. To this end, we first generate a full set of proportions of all the population in the mixture with step (increment of) 0.01 and compute all the possible v_{PR} values that each of the mixtures could produce. For step 0.01: two populations (1 tumour) give 99 proportions, three populations (2 tumours) give 4851 proportions, four populations (3 tumours) give 156849 proportions. The matrices with mixture v_{PR} values for each proportion, vary from 2x2, for two populations with deletions, to 35x35 for four populations with all events up to tetraplications. Then, we run an exhaustive approximate search over all the matrices with mixture v_{PR} values over all the proportions. The search is approximate because the estimated v_{PR} values have limited accuracy and because we consider only discrete values of proportions. In the analysis we define a match between estimated and mixture v_{PR} values if they differ <0.009. The search returns a large number of admissible mixtures that could produce the estimated v_{PR} values. This process is illustrated in Fig. 5.

To represent the admissible mixtures graphically we use ternary plots. Ternary plot allows to illustrate composition of three components using just two dimensions. The composition, represented by ratios of the three components which sum to a constant, is depicted as point inside or on the edge of an equilateral triangle. If the point is on the edges, the composition has only two components. To facilitate interpretation of the ternary plots, we also plot the grid lines that are parallel to the sides of the triangle. These gridlines indicate the directions of constant ratios of the components. Along such direction the ratio of one of the components is fixed and only the other two ratios vary. Examples of visualisation of admissible mixtures on ternary plots are shown in Figs. 5 and 6.

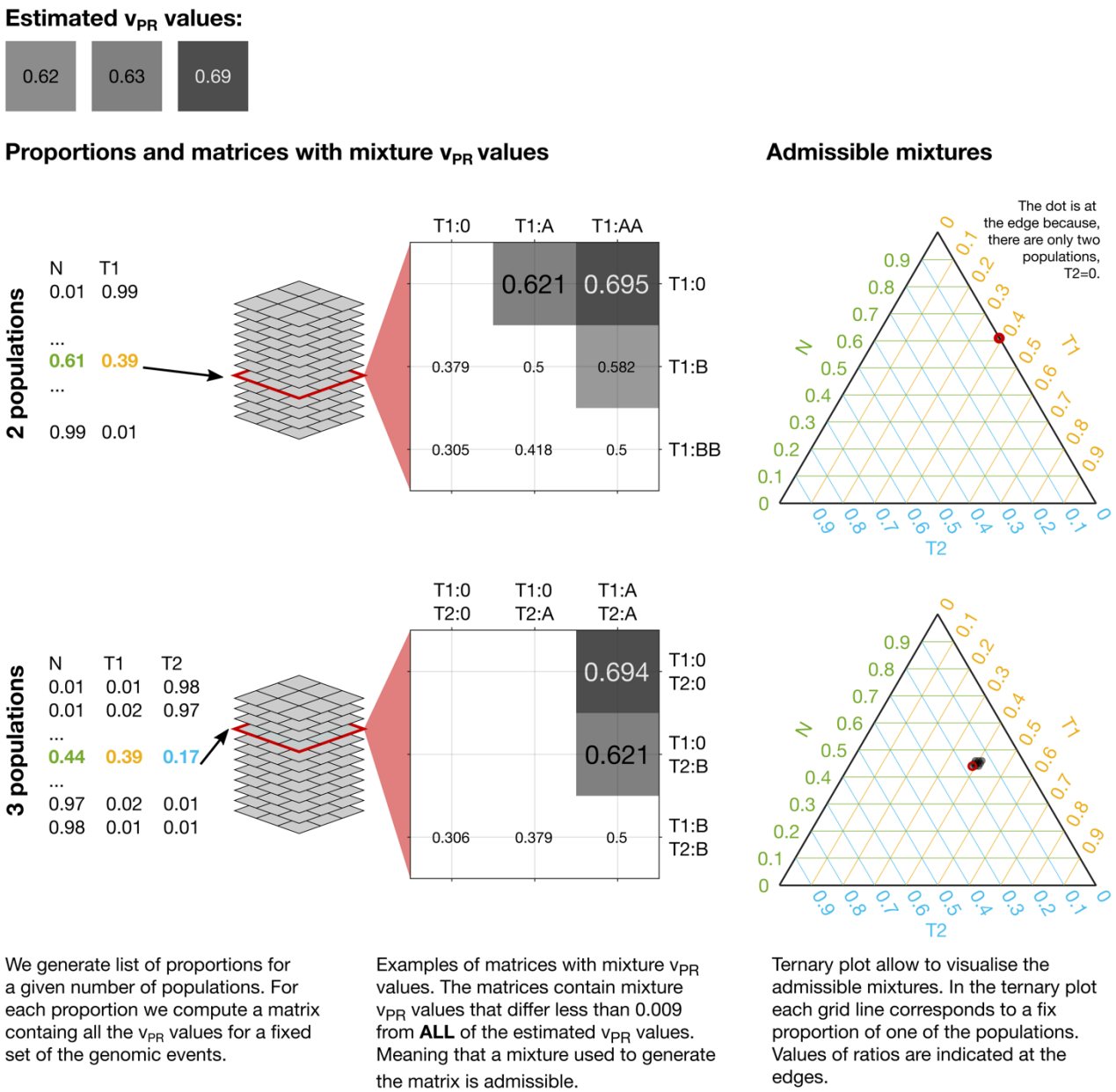


Fig. 5. Mixtures admissible by the v_{PR} values estimated from data. To uncover mixtures that could produce the three estimated v_{PR} values we perform an exhaustive approximate search of all the possible v_{PR} values produced by any mixture of the populations with a given set of genetic events. In each case we generate a full set of proportions with a given step (e.g. 0.01) and compute all the possible v_{PR} values that such a mixture could produce. In the illustrated cases: 2 populations (1 tumour) could produce the estimated v_{PR} values through a deletion (estimated v_{PR}=0.62 and v_{PR}=0.63) and via deletion of one allele and duplication of another (estimated v_{PR}=0.69); 3 populations (2 tumours) could produce the estimated v_{PR} values through a deletion in one of the tumour populations (estimated v_{PR}=0.62 and v_{PR}=0.63) and via deletion in both of the tumour populations (estimated v_{PR}=0.69). The 2 populations case admits a single mixture and the 3 populations allow 9 mixtures with similar compositions. The admissible mixtures are depicted on the ternary plots, red circle indicates solution corresponding to the presented matrix. We exclude mixture v_{PR} values that result from deletion of both the variant and reference alleles (empty fields in the matrices).

To facilitate analysis of the admissible mixtures returned by the search procedure we introduce mixture complexity. Mixture complexity is a measure that increases with number of populations as well as with variety of genetic events. From the simplest mixture of 1 normal and 1 tumour population in which only deletions are possible to a model with 1 normal and multiple tumour populations where each can have deletions, and any level of multiplications. In practice, we set the limit at 3 tumour populations and tetraplications. Mixture complexity helps to group and visualise admissible mixtures. Mixtures with higher complexity allow more possible v_{PR} values, meaning that it is easier to find the match with the estimated v_{PR} values and the number of admissible mixtures increases (see Figure 6). We, further, observe that proportion of normal population, p_N, increases with a number of clonal tumour populations included in the model mixture and that, generally, p_N stays constant with increasing variety of genetic events, for a fixed number of clonal tumour populations.

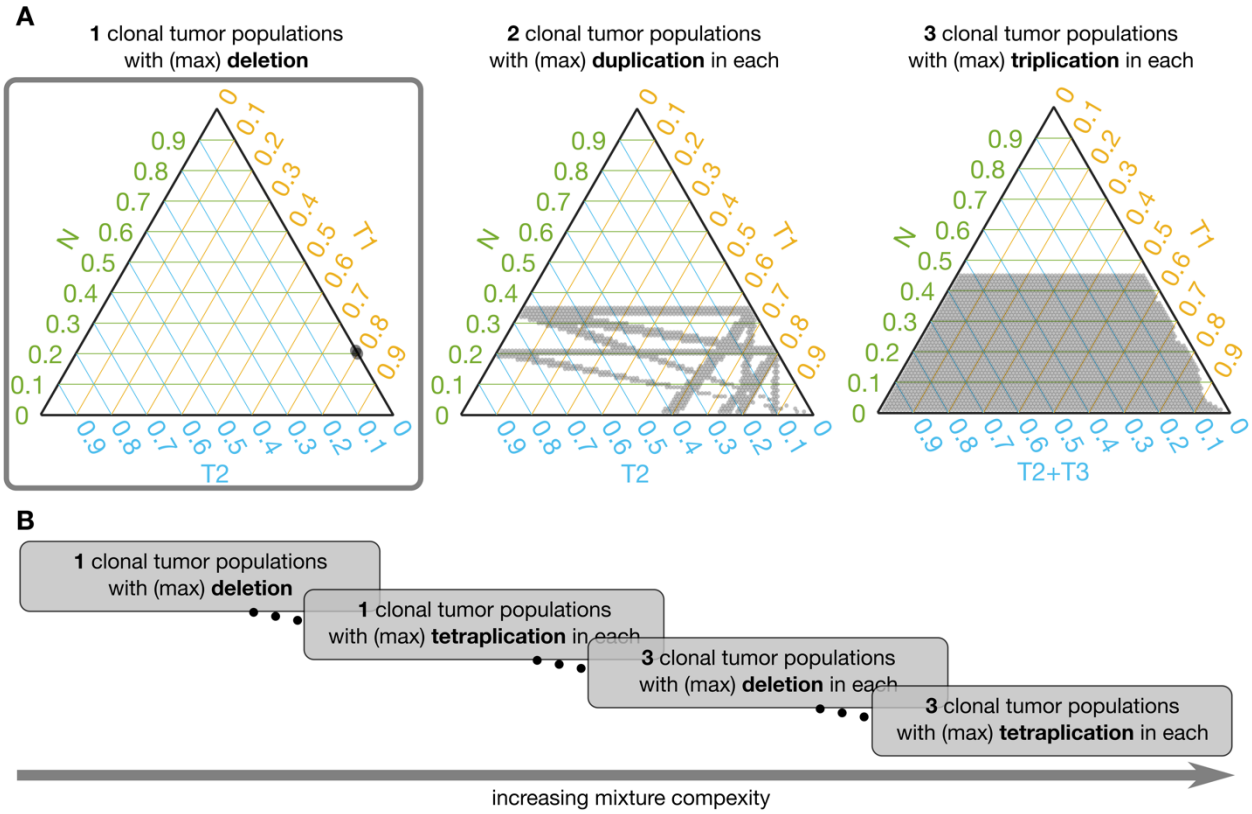


Fig. 6. Admissible mixtures for increasing mixture complexity. **A** shows admissible mixtures for 3 different mixtures with increasing complexity. The simplest mixture (mixture with the lowest number of components and the simplest set of genetic events) is show within the grey frame. On each ternary plot, the admissible mixtures are indicated by grey dots. The green axis indicates proportion of the normal population (N), the yellow axis indicates proportion of the 1st tumour population (T1), the blue axis shows proportion of the 2nd tumour population (T2) or sum of the 2nd and 3rd tumour populations (T2+T3). **B** Schematic representation of increasing complexity of the mixture models. From a mixture of 1 normal and 1 tumour population in which only deletion is possible to a model with 1 normal and 3 tumour populations and each can have deletions, du-, tri- and tetraplications.

2.2 Analysis of RNA-DNA relationships

GetAllele is readily applicable to assess RNA-DNA relationships between normal and tumour sequencing signals derived from the same sample/individual (matched datasets). As a proof of concept, we assessed matched normal and tumour exome and transcriptome sequencing data of 72 breast carcinoma (BRCA) datasets with pre-assessed copy-number and genome admixture estimation acquired through TCGA (Supplementary Table 1). For these datasets, purity and genome admixture has been assessed using at least three of the following five approaches: ESTIMATE, ABSOLUTE, LUMP, IHC, and the Consensus Purity Estimation (CPE)(Aran, et al., 2015; Carter, et al., 2012; Katkovnik, et al., 2002; Pagès, et al., 2010; Yoshihara, et al., 2013; Zheng, et al., 2014). In addition, on the same datasets we applied THetA – a popular tool for assessing CNA and admixture from sequencing data (Oesper, et al., 2013; Oesper, et al., 2014).

2.2.1 Segmentation results

Segmentation of the data, based on the tumour exome signal, resulted in 2697 windows across the 72 datasets. We excluded from further analysis 294 windows where either tumour exome or transcriptome had v_{PR}>=0.58 but their VAF distribution could not be differentiated from the model VAF distributions with v_{PR}=0.5 (p>1e-5, Kolmogorov Smirnov test, equivalent to Bonferroni FWER correction for 100000 comparisons). The 294 excluded windows correspond to 4% of the data in terms of number of base pairs in the windows and 4% of all the available data points; i.e. they are short and contain only few VAF values. In the remaining 2403 windows, we systematically examined the similarity between corresponding VAF_{TEX} (tumour exome), VAF_{TTR} (tumour transcriptome) and CNA. We obtained several distinct patterns of coordinated RNA-DNA allelic behaviour as well as correlations with CNA data.

In 60% of all analysed windows the distributions of VAF_{TEX} and VAF_{TTR} were statistically concordant (had p>1e-5, Kolmogorov Smirnov test), and in 40% they were statistically discordant (p<1e-5, Kolmogorov Smirnov test). In 2 windows VAF_{TEX} and VAF_{TTR} had the same v_{PR} but had statistically different distributions (p<1e-5, Kolmogorov Smirnov test), we consider such windows as concordant; Kolmogorov-Smirnov test is very sensitive for differences between distributions, v_{PR} fitting is more robust. In the vast majority of the discordant windows v_{PR} of the VAF_{TTR}, $v_{PR,TTR}$, was higher than v_{PR} of the VAF_{TEX}, $v_{PR,TEX}$, (only in 21 out of 959 discordant windows $v_{PR,TTR}$ was lower than $v_{PR,TEX}$).

2.2.2 Concurrence of segmentation based on DNA and RNA

We next analysed the concurrence between windows resulting from independent segmentations of the dataset based on the tumour exome and transcriptome signals in the datasets (2697 and 3605 windows, respectively, across all the samples). We first assessed chromosome-wise alignment of the start and end points of the windows. In 45% of the chromosomes both VAF_{TEX} and VAF_{TTR} signals produce a single window that contains the whole chromosome. In 33% of chromosomes both signals produced multiple windows. These windows are well aligned, with 90% of the break-points within 7% difference in terms of the number of data points in the chromosome (Q50=0.02%, Q75=2% of data points in the chromosome). The probability of observing such an alignment by chance is smaller than p=1e-5 (100,000 bootstrap samples with breaking points assigned randomly in all the individual chromosomes where both signals produced multiple windows). In 22% of the chromosome windows based on VAF_{TEX} and VAF_{TTR} the signals were positionally discordant – one signal produced a single window containing whole chromosome while the other produced multiple windows.

To compare the v_{PR} values in the 55% of chromosomes where at least one signal produced more than one window, we computed chromosome-wise mean absolute error (MAE) between the v_{PR} in two sets of windows. To account for different start and end points of the windows we interpolated the v_{PR} values (nearest neighbour interpolation) at each data point in the chromosome. We separately compared the $v_{PR,TEX}$ and $v_{PR,TTR}$ values. The alignment in terms of MAE is very good, $v_{PR,TEX}$ agreed perfectly in 11% of the chromosomes and had the percentiles of MAE equal to Q50= 0.012, Q75= 0.022 and Q97.5= 0.047, while $v_{PR,TTR}$ agreed perfectly in 8% but had slightly higher percentiles of MAE Q50= 0.019, Q75= 0.034 and Q97.5= 0.07. $v_{PR,TEX}$ and $v_{PR,TTR}$ values had MAE=0 simultaneously in 4% of the chromosomes. Probability of observing such values of MAE by chance is smaller than p=1e-3 (1000 random assignments of $v_{PR,TEX}$ and $v_{PR,TTR}$ values to windows in the 873 chromosomes where at least one signal had more than one window). It is noteworthy that MAE Q97.5<0.07 is comparable with the confidence interval of single v_{PR} estimate based on 50 VAF values. In other words, both signals in a sample (Tex and Ttr) give very similar results in terms of windows' segmentation and estimated values of the v_{PR}. Albeit, segmentation of VAF_{TTR} generates a higher number of windows. Figure 7 shows examples of concurrence between windows based on VAF_{TEX} and VAF_{TTR} signals in a positionally concordant chromosome (both signals produced multiple windows).

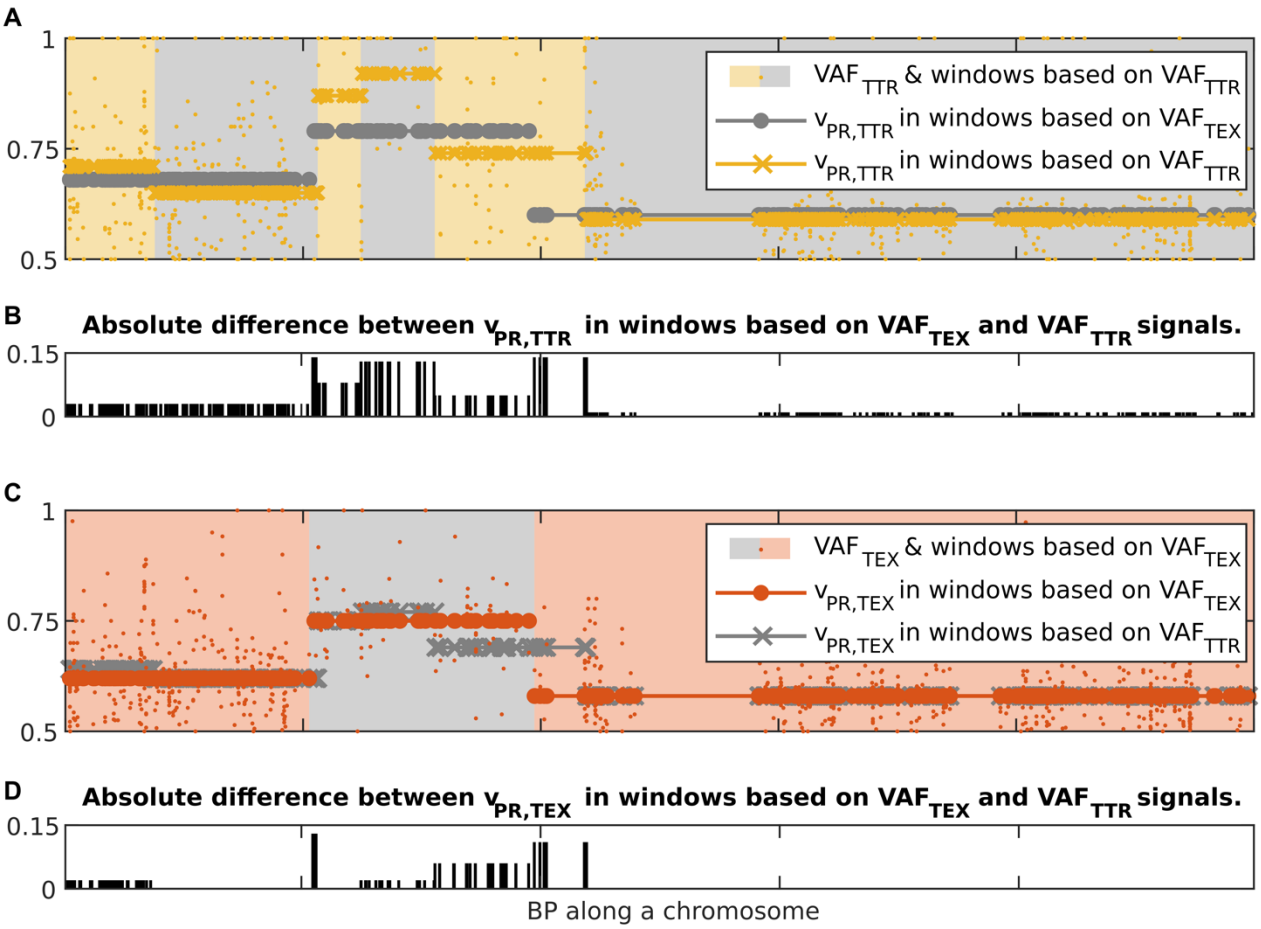


Fig. 7. Illustration of concurrence between windows resulting from independent segmentations of the dataset based on the VAF_{TEX} and VAF_{TTR} signals. **A** yellow dots, VAF_{TTR}; grey circles, $v_{PR,TTR}$ interpolated at all data points in windows based on VAF_{TEX}; yellow crosses, $v_{PR,TTR}$ interpolated at all data points in windows based on VAF_{TTR}. **B** bar plot of the absolute difference between the v_{PR} values in the two kinds of windows. **C** orange dots, VAF_{TEX}; grey crosses, $v_{PR,TEX}$ interpolated at all data points in windows based on VAF_{TEX}; orange dots $v_{PR,TEX}$ interpolated at all data points in windows based on VAF_{TTR}. **D** bar plot of the absolute difference between the v_{PR} values in the two kinds of windows.

2.2.3 Correlation between v_{PR} and CNA

Finally, we assess the correlations between v_{PR} and CNA in the individual datasets. We separately computed correlations for deletions and amplifications. In order to separate deletions and amplifications, for each data set we found CNA_{MIN}, value of the CNA in the range -0.3 to 0.3 that had the smallest corresponding $v_{PR,TEX}$. To account for observed variability of the CNA values near the CNA_{MIN}, we set the threshold for amplifications to $CNA_A = CNA_{MIN}-0.05$, and for deletions we set it to $CNA_D = CNA_{MIN}+0.05$.

For VAF_{TEX} we observed significant correlations with negative trend between $v_{PR,TEX}$ and $CNA \leq CNA_D$ in 57 datasets and with positive trend between $v_{PR,TEX}$ and $CNA \geq CNA_A$ in 39 datasets ($p_{FDR}<0.05$, Pearson's correlation with Benjamini Hochberg multiple comparison correction for 72 samples). For VAF_{TTR} we observed significant correlations with negative trend between $v_{PR,TTR}$ and $CNA \leq CNA_D$ in 62 datasets and with positive trend between $v_{PR,TTR}$ and $CNA \geq CNA_A$ in 33 datasets ($p_{FDR}<0.05$, Pearson correlation with Benjamini Hochberg correction). These correlations indicate that although $v_{PR,TEX}$ as well as $v_{PR,TTR}$ values capture information contained in CNA, they do not differentiate between positive and negative values of the CNA.

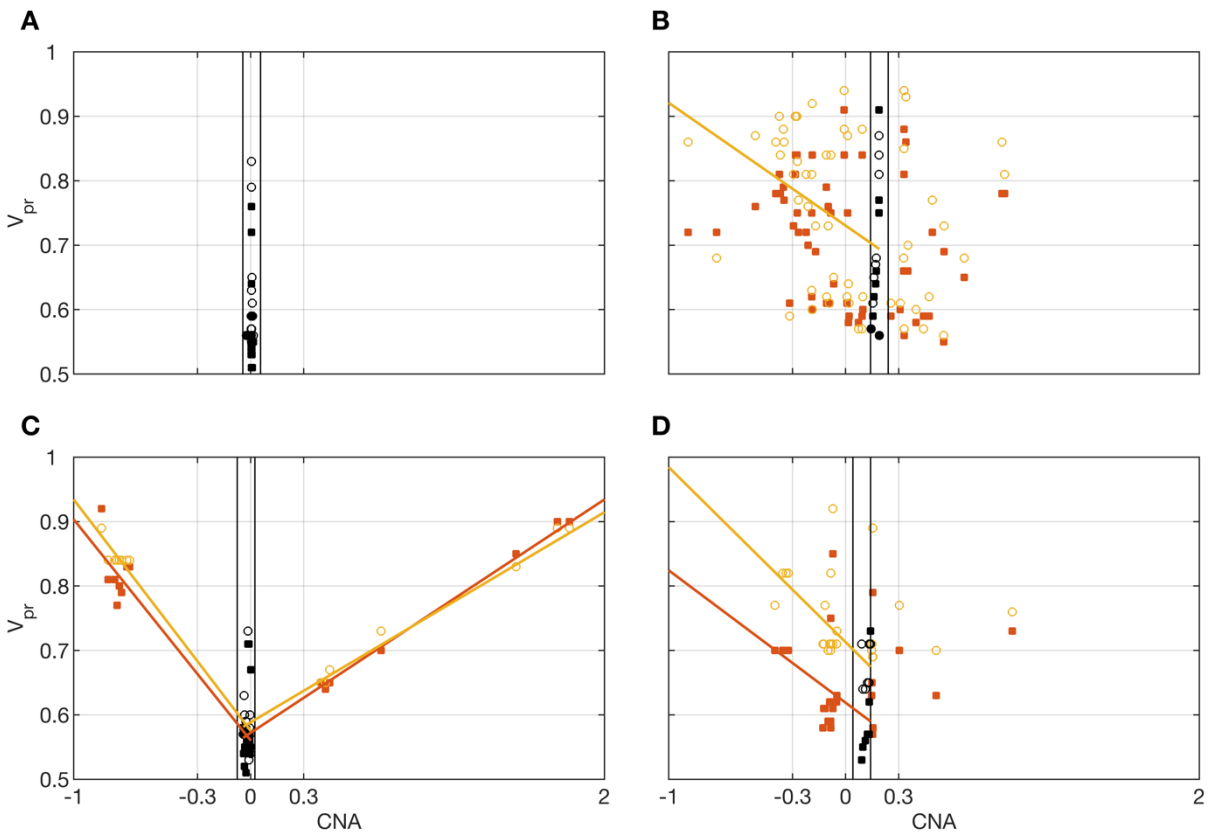


Fig. 8. Illustration of the correlations between v_{PR} and CNA. Orange squares $v_{PR,TEX}$, yellow circles $v_{PR,TTR}$. Lines, least-squares fitted trends for significant correlations (orange correlation with $v_{PR,TEX}$, yellow correlation with $v_{PR,TTR}$). Black, v_{PR} for CNA_{MIN}±0.05. Correlations for all the datasets are shown in Supplementary Figure 1. **A** all the values of CNA are close to CNA_{MIN}=0. **B** relationship between CNA and v_{PR} is noisy, only some correlations are statistically significant. **C** all the correlations are statistically significant, $v_{PR,TTR}$ values (circles) follow closely the $v_{PR,TEX}$ (squares) indicating concordance of the VAF_{TEX} and VAF_{TTR} distributions. **D** only correlations for CNA ≤ CNA_D are statistically significant.

Figure 8 shows four typical patterns of correlation between the CNA and v_{PR} values observed in the data. In Figure 8A all the values of CNA are close to CNA_{MIN}. In Figure 8B the relationship between CNA and v_{PR} is noisy, only correlation between $v_{PR,TTR}$ and CNA ≤ CNA_D are statistically significant ($r_{TEX,CNA,D}$ = -0.29, p_{FDR} = 0.063; $r_{TEX,CNA,D}$ = -0.38, p_{FDR} = 0.012; $r_{TEX,CNA,A}$ = 0.14, p_{FDR} = 0.58; $r_{TEX,CNA,A}$ = 0.19, p_{FDR} = 0.47; Pearson's correlation with Benjamini Hochberg multiple comparison correction for 72 samples). In Figure 8C all the correlations are statistically significant, $v_{PR,TTR}$ values (circles) follow closely the $v_{PR,TEX}$ (squares) indicating that in most of the windows distributions of the VAF_{TEX} and VAF_{TTR} are concordant ($r_{TEX,CNA,D}$ = -0.91, p_{FDR}<1e-10; $r_{TEX,CNA,D}$ = -0.96, p_{FDR}<1e-10; $r_{TEX,CNA,A}$ = 0.92, p_{FDR}<1e-10; $r_{TEX,CNA,A}$ = 0.95, p_{FDR}<1e-10). In Figure 8D correlations between $v_{PR,TEX}$, $v_{PR,TTR}$ and CNA ≤ CNA_D are statistically significant, but there is a big difference (with median of 0.18) between $v_{PR,TEX}$ and $v_{PR,TTR}$ values, indicating that in most of the windows the distributions of the VAF_{TEX} and VAF_{TTR} in this dataset are discordant ($r_{TEX,CNA,D}$ = -0.44, p_{FDR} = 0.047; $r_{TEX,CNA,D}$ = -0.64, p_{FDR} = 0.0017; $r_{TEX,CNA,A}$ = 0.44 p_{FDR} = 0.16; $r_{TEX,CNA,A}$ = 0.28, p_{FDR} = 0.41). In many of the datasets we observe that the $v_{PR,TTR}$ values are higher than the corresponding $v_{PR,TEX}$ values (median $v_{PR,TTR}$–$v_{PR,TEX}$ = 0.03). Correlations between v_{PR} and CNA in all datasets are shown in the Supplementary Figure 1.

2.3 v_{PR} based purity estimation

To demonstrate a practical application of estimating the admissible mixtures by means of comparing the estimated $v_{PR,TEX}$ values with the mixture v_{PR} values we next use them to estimate purity of the samples. To this end we compared the v_{PR} based purity (VBP) estimates with ESTIMATE, ABSOLUTE, LUMP, IHC, and the Consensus Purity Estimation (CPE)(Aran, et al., 2015; Carter, et al., 2012; Katkovnik, et al., 2002; Pagès, et al., 2010; Yoshihara, et al., 2013; Zheng, et al., 2014).

To obtain the VBP estimate we used $v_{PR,TEX}$ values. We, first, selected the $v_{PR,TEX}$ values that: 1. are estimated with high confidence, i.e. are based on at least 50 data points; 2. are most likely heterozygous in normal exome, i.e., have a corresponding v_{PR} value in normal exome $v_{PR,NEX}$ <0.59; 3. most likely have $v_{PR,TEX}$ >

0.5, i.e., their p-value for comparison with $v_{PR,TEX} = 0.5$ is very small $p<1e-5$ (Kolmogorov-Smirnov test). Next, we used the selected $v_{PR,TEX}$ values to find all admissible mixtures (with 1 to 3 tumour populations and allowing for all events, from deletions to tetraplications). To estimate the VBP, out of all the admissible mixtures we chose these with lowest mixture complexity and among these mixtures we take one with the highest p_N (proportion of the normal population). The VBP, percentage of tumour populations in the sample, is then given as $1-p_N$. Such approach provides rather conservative estimates of VBP (the smallest $1-p_N$). However, GetAllele can be extended to offer alternative methods of employing the admissible mixtures to estimate VBP. Development, analysis and comparison of alternative VBP estimation methods is beyond scope of the current paper.

Figure 9A shows violin plots of all considered $1-p_N$ values and (x) indicates the smallest value taken as a VBP estimate. In two of the datasets we could not estimate the purity due to lack of suitable $v_{PR,TEX}$ values. The VBP estimates shows the best agreement with ABSOLUTE method ($r=0.76$, $p<3.4e-14$, Pearson's correlation, Fig. 9B2). We suppose that this is because the ABSOLUTE method is based on copy number distributions, and our analysis (Section 2.2.3) revealed high correlations between the CNAs and v_{PR} values. Similarly to the ABSOLUTE method, VBP estimates are generally lower than the other purity estimates (ESTIMATE, LUMP, IHC, CPE); see Fig. 9B1-B5.

The approach presented in this section differs from other methods for inferring genomic mixture composition (e.g. SciClone (Miller, et al., 2014), or PyClone (Roth, et al., 2014)) in that it is based on changes of the alleles multiplicity estimated along continuous multi-SNV genomic regions. The v_{PR} values used to estimate admissible mixtures are based on windows with at least 50 VAF values which extend over millions of base pairs. In this way, the presented method is complementary to the SciClone and PyClone methods that use VAFs of somatic mutations.

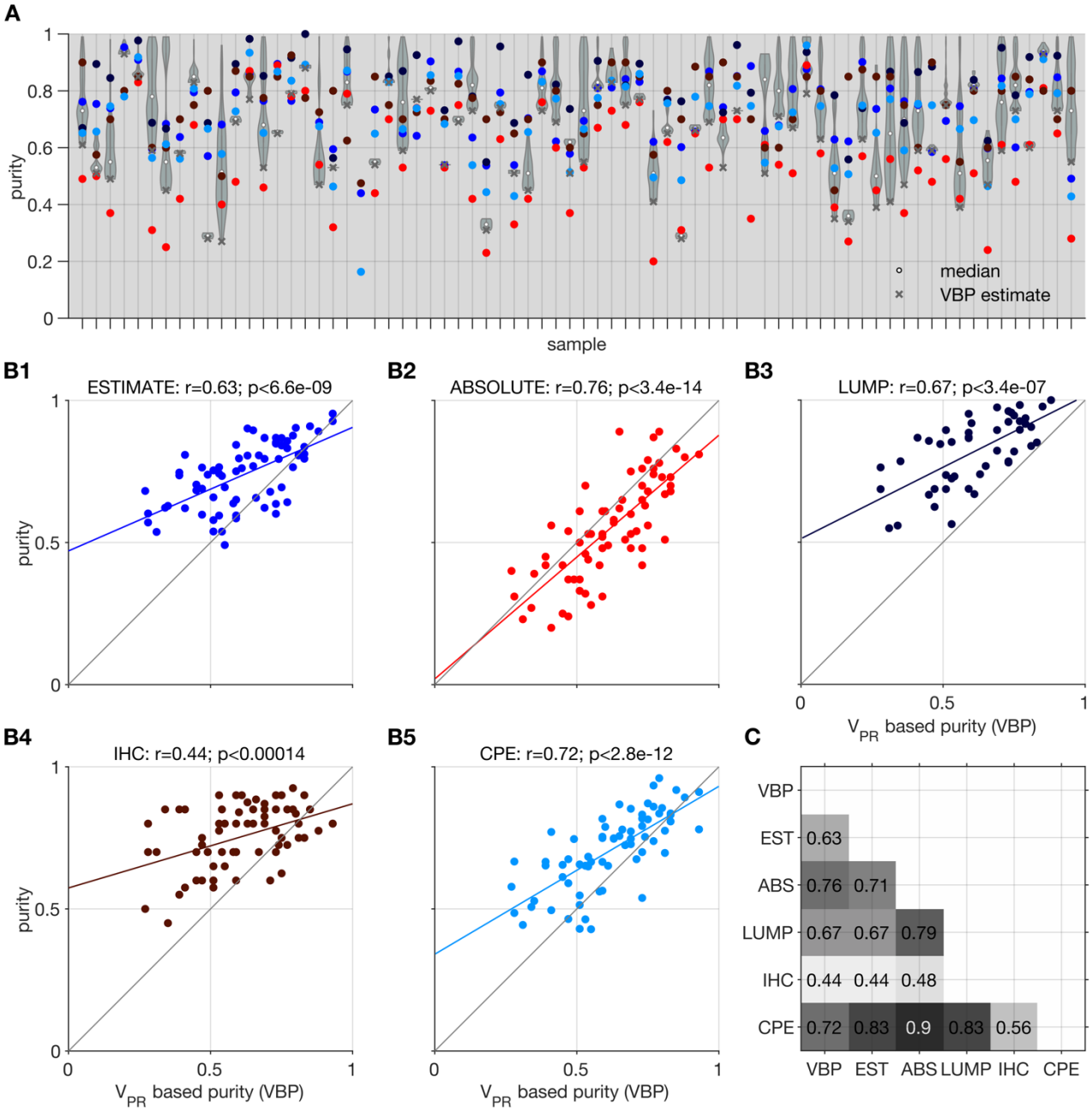


Fig. 9. Illustration of purity estimation based on model mixtures and $v_{PR,TEX}$. Comparison of the Estimate (EST), Absolute (ABS), LUMP, IHC, and the consensus purity estimate (CPE) methods with v_{PR} based purity (VBP). **A** violin plots show distributions of purity based on all the admissible proportions of the normal population (x) indicates the lowest value selected as the most conservative estimate; colours corresponding to the different methods are indicated in B1-5. **B1-5** Correlation of the VBP with individual methods; coloured line indicates best fit linear trend. **C.** Matrix showing significant p<0.05 Pearson's correlation coefficients between all tested methods.

3 Discussion

Potential for integrative analysis of RNA and DNA sequence data is facilitated by the growing availability of RNA and DNA sequencing datasets and by the technological advances now enabling simultaneous RNA and DNA sequencing from the same source (Macaulay, et al., 2016; Reuter, et al., 2016; The, et al., 2012). However, integrative RNA and DNA analyses are challenged by limited compatibility between RNA and DNA datasets and high technical variance of the sequencing-produced signals. Our approach – GeTallele – addresses the compatibility by restricting the analyses to confidently co-covered DNA and RNA regions, and the high variability - through quantification of differences between the DNA and RNA signals.

In contrast to other methods based on statistical modelling, GeTallele is based on a mechanistic model of VAF distributions that depends on the distribution of total reads (extracted from data) and the v_{PR} parameter. The brute-force simplicity and transparency of the presented methodology is one of its biggest advantages.

Additionally, in contrast to other methods, to analyse sequencing data GeTallele uses data segments that include multiple adjacent SNVs. Finally, GeTallele offers a unified pipeline to evaluate and visualise DNA-RNA whereas many of the other methods cited herein are used in conjunction with each other to achieve a similar result.

Using GeTallele, we detected several relationships between DNA-RNA allele frequencies and biological processes. First, in chromosomes affected by deletions and amplifications, VAF_{RNA} and VAF_{DNA} showed highly concordant breakpoint calls. This indicates that VAF_{RNA} alone can serve as preliminary indicator for deletions and amplifications, which can facilitate the applications of RNA-sequencing analysis on the large and constantly growing collections of transcriptome sequencing data. Second, we showcased that $v_{PR,TEX}$ can be used to estimate sample composition (in terms of proportions of normal and tumour populations) and as a result its purity. Once the mixture composition is decided by the user, GeTallele allow to further interrogate genetic events in each population at a specific data segment.

4 Conclusions

Based on our results, variant probability v_{PR} can serve as a dependable indicator to assess gene and chromosomal allele asymmetries and to aid calls of genomic events in matched sequencing RNA and DNA datasets. Methods for estimating and analysing v_{PR} values are implemented in a GeTallele toolbox. GeTallele provides a singular suit of functions for integrative analysis, statistical assessment and visualization of the observed patterns at desired resolution, including chromosome, gene, or custom genome region.

5 Methods

5.1 Samples

The GeTallele was developed using sequencing datasets from paired normal and tumour tissue obtained from 72 female patients with breast invasive carcinoma (BRCA) from The Cancer Genome Atlas (TCGA). Each of the 72 datasets contains four matched sequencing datasets: normal exome (Nex), normal transcriptome (Ntr), tumour exome (Tex), and tumour transcriptome (Ttr). In addition, we required each tumour sample to have at least three of the following five purity estimates - Estimate, Absolute, LUMP, IHC, and the consensus purity estimate (CPE), (Supplementary Table 1). Finally, each sample was required to have CNA estimation (genomic segment means based on Genome-Wide-SNPv6 hybridization array) (Aran, et al., 2015; Carter, et al., 2012; Katkovnik, et al., 2002; Pagès, et al., 2010; Yoshihara, et al., 2013; Zheng, et al., 2014).

5.2 Data processing

All datasets were generated through paired-end sequencing on an Illumina HiSeq platform. The human genome reference (hg38)-aligned sequencing reads (Binary Alignment Maps, .bams) were downloaded from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/) and processed downstream through an in-house pipeline. After variant call (Li, 2011), the RNA and DNA alignments, together with the variant lists were processed through the read count module of the package RNA2DNAlign (Movassagh, et al., 2016), to produce variant and reference sequencing read counts for all the variant positions in all four sequencing signals (normal exome, normal transcriptome, tumour exome and tumour transcriptome). Selected read count assessments were visually examined using Integrative Genomics Viewer (Thorvaldsdóttir, et al., 2013).

5.3 Data segmentation

To analyse variant allele frequencies (VAF) at genome-wide level, GeTallele first divides the VAF dataset into a set of non-overlapping windows along the chromosomes. Segmentation of the dataset into windows is based on a sequencing signal chosen out of all the available datasets in the aggregated aligned VAF dataset (one out of four in the presented analysis).

To partition the data into the windows GeTallele uses a parametric global method, which detects the breakpoints in the signals using its mean, as implemented in the Matlab function `findchangepts` (Killick, et al., 2012; Lavielle, 2005) or in R (Killick and Eckley, 2014). In each window, the VAF values of the chosen signal have a mean that is different from the mean in the adjacent windows. Sensitivity of breakpoint detection can be controlled using parameter `MinThreshold`, in the presented analysis it was set to 0.2. For segmentation and analysis (without loss of generality) we transform all the original VAF values to VAF=|VAF-0.5|+0.5.

5.4 Statistics

To test statistical significance, GeTallele uses parametric and non-parametric methods and statistical tests (Corder and Foreman, 2014; Hollander, et al., 2013). Namely, to compare distributions of the variant allele frequencies (VAF) we use Kolmogorov–Smirnov test (examples of VAF distributions are depicted in Figs. 2 and 3). To study concurrence of windows, we use permutation/ bootstrap tests. To test relations between v_{PR} and copy number alterations (CNA) we use Pearson's correlation coefficient.

To account for multiple comparisons, we set the probability for rejecting the null hypothesis at $p<1e-5$, which corresponds to Bonferroni (Dunn, 1961) family-wise error rate (FWER) correction against 100000 comparisons. We use a fixed value, rather than other approaches, to ensure better consistency and reproducibility of the results. Alternatively, we apply Benjamini and Hochberg (Benjamini and Hochberg, 1995) false discovery rate (FDR) correction with a probability of accepting false positive results $p_{FDR}<0.05$. We specify the method used in the text when reporting the results.

List of abbreviations

BRCA – breast invasive carcinoma,
CDF – cumulative distribution function,
CNA – copy number alterations,
CNA_D – copy number alterations corresponding to deletions (see section 2.2.3),
CNA_A– copy number alterations corresponding to amplifications (see section 2.2.3),
CPE – consensus purity estimate,
DNA – genome,
EMD – earth mover's distance
FWER – family-wise error rate,
FDR – false discovery rate,
MEA – mean absolute error
Nex – normal exome,
Ntr – normal transcriptome,
$r_{TEX,CNA,D}$ – Pearson's correlation coefficient between $v_{PR,TEX}$ and CNA_D
$r_{TEX,CNA,A}$ – Pearson's correlation coefficient between $v_{PR,TEX}$ and CNA_A
$r_{TTR,CNA,D}$ – Pearson's correlation coefficient between $v_{PR,TTR}$ and CNA_D
$r_{TTR,CNA,A}$ – Pearson's correlation coefficient between $v_{PR,TTR}$ and CNA_A
p_{FDR} – p-value after multiple comparisons Benjamini and Hochberg false discovery rate correction,
PDF – probability density function,
QN (e.g. Q50) – N-th percentile
RNA – transcriptome
SNV – single-nucleotide variant
TCGA – the cancer genome atlas,
Tex – tumour exome,
Ttr – tumour transcriptome,
VAF – variant allele frequency,
VAF_{TEX} – variant allele frequency in tumour exome sequence,
VAF_{TTR} – variant allele frequency in tumour transcriptome sequence,
VBP – v_{PR} based purity,
v_{PR} – variant probability,
$v_{PR,TEX}$ – variant probability estimated from tumour exome sequence,
$v_{PR,TTR}$ – variant probability estimated from tumour transcriptome sequence,

Declarations

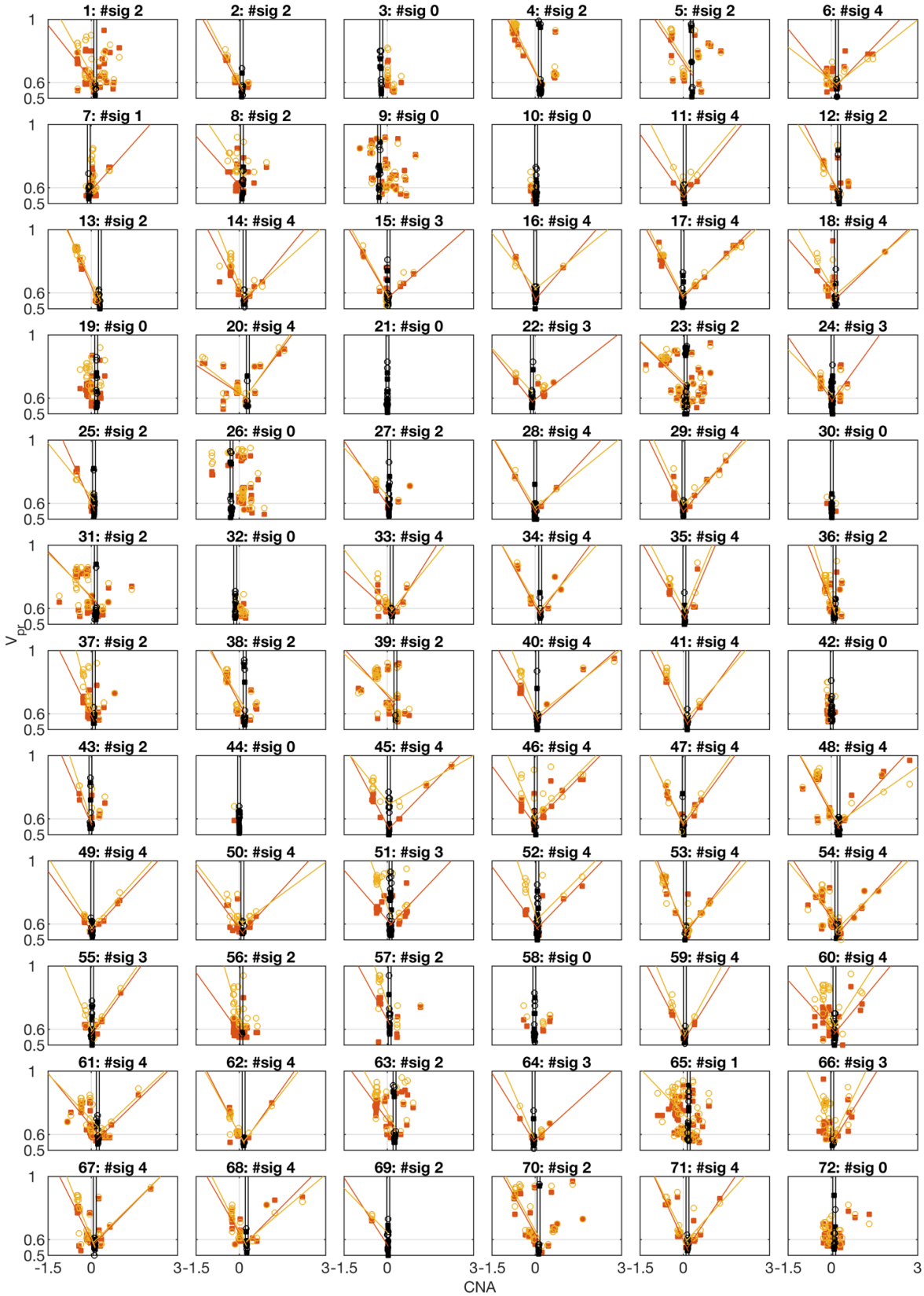
- **Ethics approval and consent to participate:** Not applicable
- **Consent for publication:** Not applicable
- **Availability of data and materials:** The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request. GeTallele is implemented as a Matlab toolbox available at: https://github.com/SlowinskiPiotr/GeTallele. We are working on R implementation of GeTallele.
- **Competing interests:** The authors declare that they have no competing interests.

- **Funding:** This work was supported by McCormick Genomic and Proteomic Center (MGPC), The George Washington University; [MGPC_PG2018 to AH]. Work of PS was generously supported by the Wellcome Trust Institutional Strategic Support Award [204909/Z/16/Z]. KTA gratefully acknowledges the financial support of the EPSRC via grant EP/N014391/1.
- **Authors' contributions:** PS, ML, PR, NA, LFS, CM, KTA, AH conception and design of the work; ML data acquisition; PS data analysis; PS, ML, PR, LFS, KTA, AH interpretation of data; PS the creation of new software used in the work; PS, ML, PR, LFS, KTA, AH have drafted the work or substantively revised it; all authors approved the submitted version. All authors agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.
- **Acknowledgements:** Not applicable

References

Aran, D., Sirota, M. and Butte, A.J. Systematic pan-cancer analysis of tumour purity. Nature communications 2015;6:8971.
Benjamini, Y. and Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met 1995;57(1):289-300.
Carter, S.L., et al. Absolute quantification of somatic DNA alterations in human cancer. Nature biotechnology 2012;30(5):413-421.
Corder, G.W. and Foreman, D.I. Nonparametric Statistics. John Wiley & Sons; 2014.
Dunn, O.J. Multiple Comparisons among Means. J Am Stat Assoc 1961;56(293):52-64.
Ferreira, E., Shaw, D.M. and Oddo, S. Identification of learning-induced changes in protein networks in the hippocampi of a mouse model of Alzheimer's disease. Translational psychiatry 2016;6(7):e849.
Ha, G., et al. Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. Genome research 2012;22(10):1995-2007.
Han, L., et al. Alternative applications for distinct RNA sequencing strategies. Briefings in bioinformatics 2015;16(4):629-639.
Hardy, G.H., et al. An introduction to the theory of numbers. Oxford ; New York: Oxford University Press; 2008.
Hollander, M., Wolfe, D.A. and Chicken, E. Nonparametric Statistical Methods. John Wiley & Sons; 2013.
Kantorovich, L.V. and Rubinstein, G.S. On a space of completely additive functions. Vestnik Leningrad. Univ 1958;13(7):52-59.
Katkovnik, V., Kgiazarian, K. and Astola, J. Adaptive window size image de-noising based on intersection of confidence intervals (ICI) rule. J Math Imaging Vis 2002;16(3):223-235.
Kendall, M.G. A new measure of rank correlation. Biometrika 1938;30(1-2):81-93.
Killick, R. and Eckley I.A. changepoint: An R Package for Changepoint Analysis. Journal of Statistical Software, 2014;58(3):1–19. http://www.jstatsoft.org/v58/i03/.
Killick, R., Fearnhead, P. and Eckley, I.A. Optimal Detection of Changepoints With a Linear Computational Cost. J Am Stat Assoc 2012;107(500):1590-1598.
Kowalski, C.J. On the Effects of Non-Normality on the Distribution of the Sample Product-Moment Correlation Coefficient. Journal of the Royal Statistical Society. Series C (Applied Statistics) 1972;21(1):1-12.
Lavielle, M. Using penalized contrasts for the change-point problem. Signal Process 2005;85(8):1501-1510.
Levina, E. and Bickel, P. The Earth Mover's distance is the Mallows distance: some insights from statistics. In, IEEE International Conference on Computer Vision
2001. p. 251-256.
Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011;27(21):2987-2993.
Li, H., et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25(16):2078-2079.
Macaulay, I.C., et al. Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq. Nature protocols 2016;11(11):2081-2103.
Miller CA, White BS, Dees ND, Griffith M, Welch JS, et al. SciClone: Inferring Clonal Architecture and Tracking the Spatial and Temporal Patterns of Tumor Evolution. PLOS Computational Biology 2014;10(8):e1003665. https://doi.org/10.1371/journal.pcbi.1003665
Morin, R.D., et al. Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. Blood 2013;122(7):1256-1265.
Movassagh, M., et al. RNA2DNAlign: nucleotide resolution allele asymmetries through quantitative assessment of RNA and DNA paired sequencing data. Nucleic acids research 2016;44(22):e161.
Newson, R. Parameters behind "nonparametric" statistics: Kendall's tau,Somers' D and median differences. Stata Journal 2002;2(1):45-64.
Oesper, L., Mahmoody, A. and Raphael, B.J. THetA: inferring intra-tumor heterogeneity from high-throughput DNA sequencing data. Genome biology 2013;14(7):R80.
Oesper, L., Satas, G. and Raphael, B.J. Quantifying tumor heterogeneity in whole-genome and whole-exome sequencing data. Bioinformatics (Oxford, England) 2014;30(24):3532-3540.
Pagès, F., et al. Immune infiltration in human tumors: a prognostic factor that should not be ignored. Oncogene 2010;29(8):1093-1102.
Reuter, J.A., et al. Simul-seq: combined DNA and RNA sequencing for whole-genome and transcriptome profiling. Nature methods 2016;13(11):953-958.
Roth, A., Khattra, J., Yap, D. et al. PyClone: statistical inference of clonal population structure in cancer. Nat Methods 2014; 11, 396–398. https://doi.org/10.1038/nmeth.2883
Shah, S.P., et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 2012;486(7403):395-399.

Shi, L., et al. Long-read sequencing and de novo assembly of a Chinese genome. Nature communications 2016;7:12065.
Shlien, A., et al. Direct Transcriptional Consequences of Somatic Mutation in Breast Cancer. Cell reports 2016;16(7):2032-2046.
The, E.P.C., et al. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57.
Thorvaldsdóttir, H., Robinson, J.T. and Mesirov, J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in bioinformatics 2013;14(2):178-192.
Yang, S., et al. An Integrated Approach for RNA-seq Data Normalization. Cancer informatics 2016;15:129-141.
Yoshihara, K., et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nature communications 2013;4:2612.
Zheng, X., et al. MethylPurify: tumor purity deconvolution and differential methylation detection from single tumor DNA methylomes. Genome biology 2014;15(8):419.



Supplementary Fig. 1. Illustration of the correlations between v_{PR} and CNA. Orange squares $v_{PR,TEX}$, yellow circles $v_{PR,TTR}$. Lines, least-squares fitted trends for significant correlations (orange correlation with $v_{PR,TEX}$, yellow correlation with $v_{PR,TTR}$). Black, v_{PR} for $CNA_{MIN}\pm0.05$. Title format Number of the dataset: #sig number of significant correlations in the dataset.

Supplementary Table 1. Datasets, signals and purity estimates.

#	TCGA BRCA datasets	EST	ABS	LUMP	IHC	CPE
1	001_Nex_BRCA_TCGA-BH-A1FC-11A_413b80f6-f6cf-4992-804a-f045e38cbe6f 001_Ntr_BRCA_TCGA-BH-A1FC-11A_086db136-f3f2-42fa-aca1-63847de6ccb9 001_Tex_BRCA_TCGA-BH-A1FC-01A_1a2187a6-aea8-4096-8c3f-208a8467cd5a 001_Ttr_BRCA_TCGA-BH-A1FC-01A_b5e2f568-e6fc-4192-a3ab-da956e5bfa4c	0.7615	0.49	0.6693	0.9	0.6517
2	002_Nex_BRCA_TCGA-BH-A0B5-11A_c724807c-d80d-4582-8238-8339397b6aec 002_Ntr_BRCA_TCGA-BH-A0B5-11A_f478930d-216a-40ec-b434-bfc3a7b2f62b 002_Tex_BRCA_TCGA-BH-A0B5-01A_803de3d6-895f-4ad1-a86c-6f72d6ea8430 002_Ttr_BRCA_TCGA-BH-A0B5-01A_37175dfe-e34e-4f97-88b1-c0ba4bd5d093	0.7539	0.5	0.8944	0.575	0.6566
3	003_Nex_BRCA_TCGA-BH-A0BJ-11A_a9988fbb-090a-4363-bf73-7505e1710623 003_Ntr_BRCA_TCGA-BH-A0BJ-11A_2ced85bc-852a-4056-ad11-2e88ec6d2d82 003_Tex_BRCA_TCGA-BH-A0BJ-01A_58ec1111-c932-49ea-9327-1c64dfc2afa6 003_Ttr_BRCA_TCGA-BH-A0BJ-01A_73442f2d-3453-42ee-b57a-86871e2e2fd9	0.7386	0.37	0.8453	0.7	0.7458
4	004_Nex_BRCA_TCGA-E2-A158-11A_58fe3067-8198-486e-b0b2-286dc4451c39 004_Ntr_BRCA_TCGA-E2-A158-11A_323eb80d-71e2-4223-b471-a83ee42e6e08 004_Tex_BRCA_TCGA-E2-A158-01A_0329fa7e-d768-4bbe-940e-36f0b9829d7c 004_Ttr_BRCA_TCGA-E2-A158-01A_9d31f395-85e7-4ad8-95a3-0cc796c4b81d	0.9534	NaN	NaN	0.8	0.7799
5	005_Nex_BRCA_TCGA-A7-A13E-11A_bd7e6f8f-7213-4ded-a8ca-3c73c7b8d918 005_Ntr_BRCA_TCGA-A7-A13E-11A_99c08ce4-6526-4982-9bc7-b9c07972bcdb 005_Tex_BRCA_TCGA-A7-A13E-01A_28b8b84b-ca69-4c6a-860c-989777b18d32 005_Ttr_BRCA_TCGA-A7-A13E-01A_148d5aec-6026-46b5-b40c-38a1198175ab	0.909	0.83	0.9772	0.85	0.9184
6	006_Nex_BRCA_TCGA-BH-A208-11A_645b786f-1942-4cce-973b-4a75956265f5 006_Ntr_BRCA_TCGA-BH-A208-11A_a6dd96f4-f194-4c8d-8757-9e8b35465a9f 006_Tex_BRCA_TCGA-BH-A208-01A_5bdbd7db-ced4-4446-9069-c44c9c1f0ae0 006_Ttr_BRCA_TCGA-BH-A208-01A_794bcf95-8e66-4f91-a49c-ab10defe73c5	0.5951	0.31	0.6877	0.6	0.5642
7	007_Nex_BRCA_TCGA-BH-A1FU-11A_db7e821b-a2b6-40e1-9fbc-c72231b703a4 007_Ntr_BRCA_TCGA-BH-A1FU-11A_c051b92b-8e11-4623-b11b-3a0d52710663 007_Tex_BRCA_TCGA-BH-A1FU-01A_cb37bb7f-8fb6-432a-a58a-f8178d5baa64 007_Ttr_BRCA_TCGA-BH-A1FU-01A_7ea95c3a-b1a6-4658-b4c2-f35f3f48394e	0.6835	0.25	0.667	0.6	0.6121
8	008_Nex_BRCA_TCGA-BH-A0AY-11A_2ecc0325-3973-48b3-b53b-bb52aea5a9bc 008_Ntr_BRCA_TCGA-BH-A0AY-11A_1b2877ac-94a0-464c-b58b-9ce2f16aff37 008_Tex_BRCA_TCGA-BH-A0AY-01A_357ccb95-03e5-49f6-ab18-38d4c8d4d820 008_Ttr_BRCA_TCGA-BH-A0AY-01A_a19a60e7-e5ca-4f66-96fe-c9add702177d	0.6376	0.42	NaN	0.7	0.5612
9	009_Nex_BRCA_TCGA-BH-A18U-11A_bf3d62cb-f3a6-45d6-b9c3-416e58f1d319 009_Ntr_BRCA_TCGA-BH-A18U-11A_9d4c1d7e-dd77-41d1-b1df-144e7afb2141 009_Tex_BRCA_TCGA-BH-A18U-01A_a80933e5-3b07-41dc-b7f0-499d63c071a9 009_Ttr_BRCA_TCGA-BH-A18U-01A_ff89e0d9-7e6c-4b6b-a1c3-f800aaa414a1	0.7949	0.68	NaN	0.75	0.8077
10	010_Nex_BRCA_TCGA-AC-A2FF-11A_714e11fb-be71-4bbd-9327-457883a07ef0 010_Ntr_BRCA_TCGA-AC-A2FF-11A_4d32c4fa-959e-41cf-b837-104290bab9fa 010_Tex_BRCA_TCGA-AC-A2FF-01A_5c6fe1fc-839c-422a-89e7-4a54dcfad6c2 010_Ttr_BRCA_TCGA-AC-A2FF-01A_37bf962c-b180-4cc2-8e0b-fde78b4f99f4	0.5705	NaN	0.6868	0.8	0.6667
11	011_Nex_BRCA_TCGA-BH-A0BQ-11A_a5bdd116-8c1b-4787-be01-4c0f96709cc5 011_Ntr_BRCA_TCGA-BH-A0BQ-11A_45e17d22-fbed-418b-97fc-7104e1deeac1 011_Tex_BRCA_TCGA-BH-A0BQ-01A_27138381-1865-4a6a-bd70-58725c92cb49 011_Ttr_BRCA_TCGA-BH-A0BQ-01A_8879454d-b803-40b6-b3d7-fbc295de9df6	0.6814	0.4	NaN	0.5	0.5779
12	012_Nex_BRCA_TCGA-BH-A0BA-11A_9dbc7f19-30bd-48fd-8d5a-ca67dc26c5b1 012_Ntr_BRCA_TCGA-BH-A0BA-11A_2cc17895-0a6e-4703-8164-7034f5c2e1a8 012_Tex_BRCA_TCGA-BH-A0BA-01A_b4c0df66-54c1-4bbf-9a3c-d2fd28d5bb4b 012_Ttr_BRCA_TCGA-BH-A0BA-01A_a9f9701c-6b4b-48ed-af83-94804fb098a8	0.7944	0.48	0.8945	0.87	0.7278
13	013_Nex_BRCA_TCGA-BH-A0B8-11A_ef67ace2-01d6-4e8b-92c7-7c4e1e5ca327 013_Ntr_BRCA_TCGA-BH-A0B8-11A_3a833d6d-75c7-4381-8cef-699c633b64e6 013_Tex_BRCA_TCGA-BH-A0B8-01A_54972439-f9da-497d-a605-24e9670021ad 013_Ttr_BRCA_TCGA-BH-A0B8-01A_9c7776d0-33df-4bd7-a720-807c650fdbc5	0.8571	0.87	0.9827	0.85	0.9342
14	014_Nex_BRCA_TCGA-BH-A0AU-11A_15483d36-ad24-4771-a991-8a8435effc6a 014_Ntr_BRCA_TCGA-BH-A0AU-11A_7f667d91-04aa-48e8-b675-9d99b64b2058 014_Tex_BRCA_TCGA-BH-A0AU-01A_e7a641f3-cc31-4319-a04b-75c42e991711 014_Ttr_BRCA_TCGA-BH-A0AU-01A_23e09239-bfc3-4c2e-b690-db940d5292f7	0.765	0.46	0.8525	0.775	0.652
15	015_Nex_BRCA_TCGA-BH-A18S-11A_9e6d6a2d-ce9e-4d44-9603-f843ffa06c63 015_Ntr_BRCA_TCGA-BH-A18S-11A_b54a0f88-21be-4c6d-a27a-1c1b8959652c 015_Tex_BRCA_TCGA-BH-A18S-01A_d4746397-9268-460a-954b-e5b5921138f9 015_Ttr_BRCA_TCGA-BH-A18S-01A_e0a3ea3a-ffce-4e30-9f42-cb047a7644a1	0.8948	0.89	NaN	0.85	0.8676

16	016_Nex_BRCA_TCGA-BH-A0HK-11A_d256dce0-d74b-4f8f-bf47-40b1b953fc7f 016_Ntr_BRCA_TCGA-BH-A0HK-11A_438650e8-0ee2-4c74-8432-88b5c8006187 016_Tex_BRCA_TCGA-BH-A0HK-01A_944b4c29-bf72-4eec-b277-badc237730de 016_Ttr_BRCA_TCGA-BH-A0HK-01A_fe04f368-0a73-4f97-9d6b-2986f9b2b052	0.7649	0.78	0.9163	0.925	0.8357
17	017_Nex_BRCA_TCGA-A7-A0D9-11A_dda70534-0d4d-4c30-9c6a-fb3c39396fb0 017_Ntr_BRCA_TCGA-A7-A0D9-11A_17cf6364-e228-4ee9-bffa-d1ad75f4152b 017_Tex_BRCA_TCGA-A7-A0D9-01A_821d7a33-77fb-496e-be9c-0552b12cbbee 017_Ttr_BRCA_TCGA-A7-A0D9-01A_c0ecd314-9d99-48ec-83f1-5a0c1ed656aa	0.8911	0.8	1	0.775	0.8921
18	018_Nex_BRCA_TCGA-BH-A0BV-11A_56dfc492-2b1f-4494-9ba9-14a70601ae21 018_Ntr_BRCA_TCGA-BH-A0BV-11A_20459115-d7be-4d04-896f-c5ff6923ec4c 018_Tex_BRCA_TCGA-BH-A0BV-01A_beb9e4cf-1f76-4a26-acee-e88d0936e60b 018_Ttr_BRCA_TCGA-BH-A0BV-01A_d037d3c2-e316-473d-9970-d4fb43615d95	0.6895	0.54	NaN	0.725	0.6749
19	019_Nex_BRCA_TCGA-E2-A1LH-11A_61558dd3-8f6c-4f70-8717-7676580fa5a7 019_Ntr_BRCA_TCGA-E2-A1LH-11A_c7e02b93-465f-47da-81d7-ec9a8cb1e52b 019_Tex_BRCA_TCGA-E2-A1LH-01A_f54770bb-5dd0-48cf-ac5a-3f023a6aef95 019_Ttr_BRCA_TCGA-E2-A1LH-01A_169c390c-a211-4db0-a983-9bf5d6eee16e	0.5948	0.32	0.5637	0.8	0.4633
20	020_Nex_BRCA_TCGA-BH-A0DD-11A_e9fe9b97-f7c7-40dc-ae31-17bb15c9fd8b 020_Ntr_BRCA_TCGA-BH-A0DD-11A_5482cdd0-3698-455b-97c1-b10c69d67ae9 020_Tex_BRCA_TCGA-BH-A0DD-01A_99ca9706-f2bf-430b-9b23-e0947c0f8593 020_Ttr_BRCA_TCGA-BH-A0DD-01A_90cbc532-1ca8-46d6-977c-72b6d01e9c34	0.8677	0.79	0.9459	0.625	0.8714
21	021_Nex_BRCA_TCGA-BH-A0H5-11A_adfb1a86-fbb1-4b71-9c04-f99399f20d70 021_Ntr_BRCA_TCGA-BH-A0H5-11A_896d76a1-bae8-495a-9e12-e82e16bd8b16 021_Tex_BRCA_TCGA-BH-A0H5-01A_6cc3c90e-c77c-4609-ada5-9b78c659dc34 021_Ttr_BRCA_TCGA-BH-A0H5-01A_778c9326-998d-4081-b148-0eede2b94e29	0.4399	NaN	NaN	0.475	0.1632
22	022_Nex_BRCA_TCGA-A7-A0DB-11A_91081819-79c8-4de6-bfdb-742df760c08b 022_Ntr_BRCA_TCGA-A7-A0DB-11A_a8ed2ec3-0285-4028-9698-710a148ce11b 022_Tex_BRCA_TCGA-A7-A0DB-01A_37a9daca-9d53-4ec4-8de2-dc2c140a5d8f 022_Ttr_BRCA_TCGA-A7-A0DB-01A_1f62e969-d05d-4a4d-a163-cb06e4958f71	0.7341	0.44	NaN	0.85	0.6494
23	023_Nex_BRCA_TCGA-BH-A1FN-11A_e1c0d95f-949c-4cec-9cf8-f91c3b90b8d9 023_Ntr_BRCA_TCGA-BH-A1FN-11A_d5e5f3c9-4129-4c92-87f3-6f86577a7584 023_Tex_BRCA_TCGA-BH-A1FN-01A_8d715491-6943-4d58-92f6-88cce7b463e2 023_Ttr_BRCA_TCGA-BH-A1FN-01A_8e7dc738-8a8f-45b2-bd82-4125a07d7373	0.8367	0.7	0.8509	0.75	0.8313
24	024_Nex_BRCA_TCGA-BH-A0AZ-11A_9bbae9a0-9f12-48cf-9aa7-d070c6627ea5 024_Ntr_BRCA_TCGA-BH-A0AZ-11A_693bf8e4-b266-4b58-b812-f579179efb65 024_Tex_BRCA_TCGA-BH-A0AZ-01A_664528b7-b511-4627-8464-0702263434c5 024_Ttr_BRCA_TCGA-BH-A0AZ-01A_07f377f4-0bd1-4647-bf06-ff6ed553c44a	0.6505	0.53	0.8696	0.7	0.6655
25	025_Nex_BRCA_TCGA-BH-A0HA-11A_c61bb1ab-688f-4d58-8388-60ae77c28840 025_Ntr_BRCA_TCGA-BH-A0HA-11A_09e07a68-a443-4c16-a0de-78cd8aea59c0 025_Tex_BRCA_TCGA-BH-A0HA-01A_2c144eba-6490-4d64-9446-085d6edc8308 025_Ttr_BRCA_TCGA-BH-A0HA-01A_6d483def-2d91-4afc-991a-4a29804a6f3a	0.6418	0.74	0.9258	0.725	0.7386
26	026_Nex_BRCA_TCGA-A7-A0CE-11A_eee8d4d0-d524-47f5-b076-6ad6216de1a3 026_Ntr_BRCA_TCGA-A7-A0CE-11A_548cad87-ec95-47e2-890e-7c8284ea5b88 026_Tex_BRCA_TCGA-A7-A0CE-01A_4288da4e-7e77-434b-a092-9450b0cb7833 026_Ttr_BRCA_TCGA-A7-A0CE-01A_14201682-0c8d-49c7-a5e1-7026e1a07b69	0.9035	0.73	NaN	0.835	0.8551
27	027_Nex_BRCA_TCGA-BH-A0DK-11A_3f4400a1-84ab-4198-b9a1-67b2ffc5ef36 027_Ntr_BRCA_TCGA-BH-A0DK-11A_ae67044f-62c9-405f-bfc1-f0b8f1bc66d3 027_Tex_BRCA_TCGA-BH-A0DK-01A_e3e2053a-3ca2-4527-9b94-209def68dcc3 027_Ttr_BRCA_TCGA-BH-A0DK-01A_a3df35ec-a8d2-44ad-8ba6-eaba504261e0	0.5384	0.53	0.7316	0.7	0.6837
28	028_Nex_BRCA_TCGA-BH-A0E1-11A_f6fed4ed-a853-40aa-bf7b-e627efd402d6 028_Ntr_BRCA_TCGA-BH-A0E1-11A_52441de4-e26b-42b3-b061-94907c049501 028_Tex_BRCA_TCGA-BH-A0E1-01A_3c7e6a59-08b8-4903-932a-99946a96b746 028_Ttr_BRCA_TCGA-BH-A0E1-01A_f412f8d8-9e35-41d9-b44a-131186cb4bb0	0.8676	0.75	0.9742	0.825	0.8524
29	029_Nex_BRCA_TCGA-BH-A0DG-11A_c99b1fb3-17e3-4472-86ee-7fda358a92c2 029_Ntr_BRCA_TCGA-BH-A0DG-11A_bfdaf242-1e97-450d-9983-2cbb4e99305d 029_Tex_BRCA_TCGA-BH-A0DG-01A_721e2f71-60ae-4d63-9f05-113bce56c672 029_Ttr_BRCA_TCGA-BH-A0DG-01A_865afd6b-84a7-4dde-aa23-0b925c0b9d50	0.6358	0.42	0.7804	0.775	0.5386
30	030_Nex_BRCA_TCGA-AC-A2FB-11A_552279ea-d7b1-496d-8170-ca30f5b62b5a 030_Ntr_BRCA_TCGA-AC-A2FB-11A_56cd7da0-2c47-4986-91ce-07db2bb87369 030_Tex_BRCA_TCGA-AC-A2FB-01A_de000c35-8bf4-470a-9656-1b5da0deebe6 030_Ttr_BRCA_TCGA-AC-A2FB-01A_35aa5078-e07f-4a0f-84c1-01a0e566e97c	0.5372	0.23	0.5493	0.7	0.4436

31	031_Nex_BRCA_TCGA-BH-A0H7-11A_abfca562-d328-40d2-83bb-e584123b0f28 031_Ntr_BRCA_TCGA-BH-A0H7-11A_d969d9b2-9d8b-4594-95d4-87e6ce1236fc 031_Tex_BRCA_TCGA-BH-A0H7-01A_e8daad78-39fc-4835-b1c4-8807653d9c9a 031_Ttr_BRCA_TCGA-BH-A0H7-01A_0d37f87a-760a-472a-acba-bbc255422fbe	0.7939	0.63	0.9561	0.725	0.7534
32	032_Nex_BRCA_TCGA-BH-A1EU-11A_38e87966-9605-4454-a4d1-28f96b7689f7 032_Ntr_BRCA_TCGA-BH-A1EU-11A_3b00c121-17f2-461e-8873-08d15c9ec9f4 032_Tex_BRCA_TCGA-BH-A1EU-01A_4bccbb0f-2641-44df-b89a-42f020b4c08f 032_Ttr_BRCA_TCGA-BH-A1EU-01A_86e3dba1-48fb-44cc-b046-8bc35963ce99	0.5387	0.33	0.6869	0.65	0.4299
33	033_Nex_BRCA_TCGA-BH-A0DP-11A_27543260-52ac-444b-8214-e62dca2cc8fe 033_Ntr_BRCA_TCGA-BH-A0DP-11A_30f4e5d8-a13d-4ef2-88e0-a01e07c2e142 033_Tex_BRCA_TCGA-BH-A0DP-01A_0326975a-2e56-404a-8776-92c5c5678853 033_Ttr_BRCA_TCGA-BH-A0DP-01A_7ff8a7a0-5235-4de0-bb9f-b811230b5bda	0.7037	0.42	0.8565	0.7	0.655
34	034_Nex_BRCA_TCGA-BH-A18N-11A_6c8aae77-5f43-41ec-a139-81f3ba02f6ea 034_Ntr_BRCA_TCGA-BH-A18N-11A_0738f1b2-aa50-4921-82e1-d3614b40f98d 034_Tex_BRCA_TCGA-BH-A18N-01A_b6f89799-9070-4fbc-b10c-53cbe515eccc 034_Ttr_BRCA_TCGA-BH-A18N-01A_4b7b8eb8-d939-411c-be5e-cf41a5521963	0.8685	0.76	NaN	0.9	0.832
35	035_Nex_BRCA_TCGA-BH-A0BC-11A_6359db46-f8dd-4dc8-a3a9-8725d8f6958a 035_Ntr_BRCA_TCGA-BH-A0BC-11A_2cb50d4a-d6df-4b64-acfb-7a7db5ddd1de 035_Tex_BRCA_TCGA-BH-A0BC-01A_73b9208d-336c-4990-a27d-0164a77dd165 035_Ttr_BRCA_TCGA-BH-A0BC-01A_92e26b53-f540-428a-a3c5-848a36b31171	0.6221	0.6	0.8221	0.8	0.7727
36	036_Nex_BRCA_TCGA-BH-A0BZ-11A_2b4e3d99-07cd-4b06-ad97-82a19ac0eb5d 036_Ntr_BRCA_TCGA-BH-A0BZ-11A_3aa16a4b-4e35-4530-84fb-0cb204290b08 036_Tex_BRCA_TCGA-BH-A0BZ-01A_74414845-839f-4885-b13d-3f2e17781f84 036_Ttr_BRCA_TCGA-BH-A0BZ-01A_efefcc2f-72e9-4634-b943-d26083e1a312	0.5788	0.37	0.7363	0.6	0.5138
37	037_Nex_BRCA_TCGA-BH-A0DL-11A_2d495f9c-4ffa-4169-b583-6786612e9606 037_Ntr_BRCA_TCGA-BH-A0DL-11A_bd8b100a-8391-4046-847f-c3fdd3830eeb 037_Tex_BRCA_TCGA-BH-A0DL-01A_dfd355e4-478a-47cb-9aab-8ce22b6f936c 037_Ttr_BRCA_TCGA-BH-A0DL-01A_11d77ef2-b3f9-4af9-8490-71f9a8c599e0	0.6944	0.53	NaN	0.65	0.6655
38	038_Nex_BRCA_TCGA-BH-A0BT-11A_32430467-5215-4738-86a3-5bbe11fbba86 038_Ntr_BRCA_TCGA-BH-A0BT-11A_cbef4196-5f3b-40d9-b26f-5b2bb82fbe9b 038_Tex_BRCA_TCGA-BH-A0BT-01A_e78b9962-7bc2-4238-806a-5933ac07de99 038_Ttr_BRCA_TCGA-BH-A0BT-01A_aae75165-efa0-46b3-8a8d-82dc7d82aecd	0.8096	0.67	0.9054	0.75	0.7751
39	039_Nex_BRCA_TCGA-BH-A18Q-11A_b58d4f69-a4ea-489b-9d25-e5cfdc465adb 039_Ntr_BRCA_TCGA-BH-A18Q-11A_76575097-374b-4fb2-8054-2a31b4204165 039_Tex_BRCA_TCGA-BH-A18Q-01A_1f2c90ef-a05d-494c-9232-e705691f46b9 039_Ttr_BRCA_TCGA-BH-A18Q-01A_f0173e28-7fe4-411f-a187-57fd94a7935a	0.8117	0.73	NaN	0.9	0.836
40	040_Nex_BRCA_TCGA-E2-A1LB-11A_e2d7a695-b0bf-4432-8f98-1843bb49efba 040_Ntr_BRCA_TCGA-E2-A1LB-11A_6eb518ab-f174-45ae-8d65-74086ecb1125 040_Tex_BRCA_TCGA-E2-A1LB-01A_3ddbc444-ee1b-43be-bab5-b0f67d5eb339 040_Ttr_BRCA_TCGA-E2-A1LB-01A_d7d566a0-b6d0-4a4f-8211-9309b27b0ade	0.8415	0.68	0.8192	0.9	0.815
41	041_Nex_BRCA_TCGA-BH-A0DH-11A_a7a7e0f6-100f-4145-9599-693e6c14e903 041_Ntr_BRCA_TCGA-BH-A0DH-11A_5a0374e5-cee9-4952-9df0-4ff125196478 041_Tex_BRCA_TCGA-BH-A0DH-01A_eb680f8c-4ba1-45ef-8b94-e58b68922f2f 041_Ttr_BRCA_TCGA-BH-A0DH-01A_71a3c27c-0982-4da6-b260-cf16a4868a19	0.8317	0.76	0.8955	0.85	0.8597
42	042_Nex_BRCA_TCGA-BH-A0B7-11A_d9aca915-ea30-4939-af59-edaef8872396 042_Ntr_BRCA_TCGA-BH-A0B7-11A_8db8b247-05b8-46ca-8791-ecf846da2c7f 042_Tex_BRCA_TCGA-BH-A0B7-01A_e3b9eb8a-93f3-4668-a54b-fa8b15be5667 042_Ttr_BRCA_TCGA-BH-A0B7-01A_0fdae4ee-ca68-4ba4-ba58-76058409b02f	0.6207	0.2	NaN	0.575	0.4954
43	043_Nex_BRCA_TCGA-E2-A15I-11A_36024763-f828-4496-8fdc-46d5c3de569b 043_Ntr_BRCA_TCGA-E2-A15I-11A_ffa9ace8-9253-4775-9ad2-2a8a50c0f9c9 043_Tex_BRCA_TCGA-E2-A15I-01A_8c627466-eb99-4a7e-87e6-314ae8ed32a1 043_Ttr_BRCA_TCGA-E2-A15I-01A_3a4e3785-fb2e-4ffc-9644-91c78a9a9ebe	0.7689	0.62	0.768	0.8	0.7565
44	044_Nex_BRCA_TCGA-BH-A0DV-11A_79a92eab-c87c-4209-819e-193d653c0df6 044_Ntr_BRCA_TCGA-BH-A0DV-11A_e87e7e3e-9059-47cf-9f45-8959250b037f 044_Tex_BRCA_TCGA-BH-A0DV-01A_105290eb-b626-4318-9b8a-42f477e2cec6 044_Ttr_BRCA_TCGA-BH-A0DV-01A_7bad3f4c-6065-4245-8119-c25596f38829	0.6021	0.31	0.7631	0.7	0.4856
45	045_Nex_BRCA_TCGA-BH-A0DZ-11A_aebf04d4-4a1b-4a50-b5f1-0f9e2c273121 045_Ntr_BRCA_TCGA-BH-A0DZ-11A_80b4d43d-9e7d-4ab8-b05a-0eb51faa9d12 045_Tex_BRCA_TCGA-BH-A0DZ-01A_4e7d62f5-4be9-4b9c-9b7c-aec4567dded2 045_Ttr_BRCA_TCGA-BH-A0DZ-01A_8b1982a0-315c-47e1-8de0-a1e5ec51dd74	0.6572	0.65	NaN	0.885	0.7792
46	046_Nex_BRCA_TCGA-BH-A18R-11A_b32b2067-a79e-42c5-ae78-135c845253fe 046_Ntr_BRCA_TCGA-BH-A18R-11A_f82099ae-9d74-44d8-ba5b-cd10eeb09807	0.8685	0.53	NaN	0.85	0.7466

	046_Tex_BRCA_TCGA-BH-A18R-01A_c518bc34-50dc-4265-824f-a954e4d19f0b 046_Ttr_BRCA_TCGA-BH-A18R-01A_fc65ff2e-9808-4c1e-a16b-8285fd0d27df					
47	047_Nex_BRCA_TCGA-E2-A15K-11A_6299f114-932a-42c0-8cab-bebb12c996fc 047_Ntr_BRCA_TCGA-E2-A15K-11A_c2ab9488-d9a4-479c-b9e2-6f9b0cdbaacb 047_Tex_BRCA_TCGA-E2-A15K-01A_80019ec7-b0d8-4573-b5d6-a5d9f2745ab2 047_Ttr_BRCA_TCGA-E2-A15K-01A_7e3a600b-edd8-428e-b88d-af4c63dcaad9	0.7425	0.7	0.724	0.9	NaN
48	048_Nex_BRCA_TCGA-BH-A1EN-11A_a6259119-3d8a-4749-a517-c675efbc8215 048_Ntr_BRCA_TCGA-BH-A1EN-11A_488f1b69-c2a3-429b-a972-31edfd615a67 048_Tex_BRCA_TCGA-BH-A1EN-01A_72e4cb26-911c-4804-9e4f-ed5b51024cd1 048_Ttr_BRCA_TCGA-BH-A1EN-01A_96360e75-26b6-4647-b974-9e31ae6de00c	0.8518	0.7	0.9612	0.85	0.8179
49	049_Nex_BRCA_TCGA-BH-A0H9-11A_d8b452e5-010a-4fec-80a4-770a5a492090 049_Ntr_BRCA_TCGA-BH-A0H9-11A_1337ceba-db77-4b31-ac20-1c6a8bb5f546 049_Tex_BRCA_TCGA-BH-A0H9-01A_ac4899fe-f56d-4b98-9a54-73ffd0c90652 049_Ttr_BRCA_TCGA-BH-A0H9-01A_a97d281f-235f-481b-b26b-169b96e0e65f	0.8872	0.35	0.7934	0.7	0.747
50	050_Nex_BRCA_TCGA-E2-A153-11A_1c3f2e11-952a-4e47-a8b9-25f4fe4bf205 050_Ntr_BRCA_TCGA-E2-A153-11A_bd0ab51c-c114-40e3-a6a1-4b8f576a41d3 050_Tex_BRCA_TCGA-E2-A153-01A_85258fb2-26ab-4a66-b8a5-5a58bf9275e0 050_Ttr_BRCA_TCGA-E2-A153-01A_091a54cd-e3b3-4af2-828f-a80e64504f5e	0.6589	0.61	NaN	0.6	0.5474
51	051_Nex_BRCA_TCGA-BH-A0BW-11A_ab130f7f-4070-436e-ac7b-c1b7aecb9dc6 051_Ntr_BRCA_TCGA-BH-A0BW-11A_2581c95b-1b57-4407-bbc4-a65c89bfd136 051_Tex_BRCA_TCGA-BH-A0BW-01A_7661179f-df6c-4a57-adca-224b62d98348 051_Ttr_BRCA_TCGA-BH-A0BW-01A_a15d171e-ac82-4712-94a9-b4799e7b2915	0.6784	0.54	NaN	0.6	0.6749
52	052_Nex_BRCA_TCGA-BH-A18J-11A_3aa5b173-17b2-425f-b5c7-395614d6bfa2 052_Ntr_BRCA_TCGA-BH-A18J-11A_307eb339-a781-45cc-9597-da0be7e5438a 052_Tex_BRCA_TCGA-BH-A18J-01A_f639a485-8ebb-4dcc-9f2e-a8d7ad05564f 052_Ttr_BRCA_TCGA-BH-A18J-01A_0e985713-0492-4191-918b-fef6c23389b1	0.8066	0.51	NaN	0.7	0.7243
53	053_Nex_BRCA_TCGA-BH-A204-11A_98e5c1d8-5c14-4416-8437-31d0098dd341 053_Ntr_BRCA_TCGA-BH-A204-11A_2afdc0cf-2723-42dd-89f7-fa03c6ba218c 053_Tex_BRCA_TCGA-BH-A204-01A_970600ce-9486-4641-8555-533132f7a414 053_Ttr_BRCA_TCGA-BH-A204-01A_893fcf87-8baa-424e-8866-bcf8cfa26cf9	0.8763	0.89	0.9364	0.85	0.9601
54	054_Nex_BRCA_TCGA-BH-A18K-11A_502ee86f-829e-4b6e-a8f1-be082c445310 054_Ntr_BRCA_TCGA-BH-A18K-11A_375bcdd8-8628-4047-948a-fa98bfa3dba5 054_Tex_BRCA_TCGA-BH-A18K-01A_7e8c2ea7-04ce-47c5-b848-229f96563015 054_Ttr_BRCA_TCGA-BH-A18K-01A_12473f59-359c-4306-ade3-2156e458cd05	0.8058	0.58	NaN	0.8	0.7468
55	055_Nex_BRCA_TCGA-BH-A0C3-11A_9fedd2f4-d2c8-4d24-987b-69edf55e15f1 055_Ntr_BRCA_TCGA-BH-A0C3-11A_a49fa48d-efc4-4b99-a42e-7019236af6c8 055_Tex_BRCA_TCGA-BH-A0C3-01A_866d9cb0-a299-46c1-a787-2e73fd758fbe 055_Ttr_BRCA_TCGA-BH-A0C3-01A_164f86df-dec9-44ef-b1c7-2ee5d33617be	0.6289	0.39	0.7847	0.45	0.528
56	056_Nex_BRCA_TCGA-BH-A0C0-11A_9778035c-19ff-4a89-bba2-fa83e51d9add 056_Ntr_BRCA_TCGA-BH-A0C0-11A_52a72824-0b41-4b8e-86f0-41ee5e00c989 056_Tex_BRCA_TCGA-BH-A0C0-01A_568a2363-b7e5-48f6-9242-328950eebf39 056_Ttr_BRCA_TCGA-BH-A0C0-01A_f5fe9655-f5b4-413b-882c-43b872e4ec23	0.6223	0.27	0.5586	0.85	0.5068
57	057_Nex_BRCA_TCGA-BH-A0E0-11A_59fa57c8-7435-4499-bd09-bf969596c18d 057_Ntr_BRCA_TCGA-BH-A0E0-11A_a3e5f7bd-3ab0-4ea6-9de1-742de1a2ed78 057_Tex_BRCA_TCGA-BH-A0E0-01A_72436fcd-21fd-46bd-bd36-7f514edb51de 057_Ttr_BRCA_TCGA-BH-A0E0-01A_b98d2a16-974c-4728-9648-81dc4314f225	0.9013	0.57	0.7391	0.875	0.7486
58	058_Nex_BRCA_TCGA-BH-A18M-11A_8e0cb775-9fc9-4001-973c-ef6cec2a38b6 058_Ntr_BRCA_TCGA-BH-A18M-11A_3e02aa37-30ee-4663-bba1-280e5127f302 058_Tex_BRCA_TCGA-BH-A18M-01A_69ccc418-264c-4e8e-a034-39607c07fa59 058_Ttr_BRCA_TCGA-BH-A18M-01A_aea8fff8-dbc9-4a1b-9a3a-ca1882432c57	0.7363	0.45	NaN	0.85	0.6525
59	059_Nex_BRCA_TCGA-E2-A1BC-11A_45b8b995-c477-4358-8050-6d41c267b467 059_Ntr_BRCA_TCGA-E2-A1BC-11A_205007dd-4bc1-4e1f-9fdb-115b0e7c9836 059_Tex_BRCA_TCGA-E2-A1BC-01A_f969cce4-0fcd-47bb-91f1-37ba0f314994 059_Ttr_BRCA_TCGA-E2-A1BC-01A_ea3bd91d-520b-4198-b011-a0f578eadc3e	0.8083	0.56	0.8676	0.85	0.7707
60	060_Nex_BRCA_TCGA-BH-A203-11A_f08939ed-a218-4688-b419-a91333d0267b 060_Ntr_BRCA_TCGA-BH-A203-11A_8c2d82aa-0b36-488f-8cf1-f82795b831c5 060_Tex_BRCA_TCGA-BH-A203-01A_55a9b84d-ca9f-402c-8d93-15aef2fde988 060_Ttr_BRCA_TCGA-BH-A203-01A_8986bdd8-2be5-41ed-9596-59ca1f95e1c0	0.7637	0.37	0.6872	0.75	0.5905
61	061_Nex_BRCA_TCGA-BH-A1F2-11A_d6eb2d94-1234-46b3-9403-958f3b340fd0 061_Ntr_BRCA_TCGA-BH-A1F2-11A_210014fa-f161-4799-a1c0-6f93d4b631f6 061_Tex_BRCA_TCGA-BH-A1F2-01A_91aeda5a-ed5a-4175-b19e-408219b980fc 061_Ttr_BRCA_TCGA-BH-A1F2-01A_ceb5a503-e107-4721-9283-714406cdd914	0.7507	0.52	0.8661	0.6	0.754

62	062_Nex_BRCA_TCGA-BH-A1EO-11A_90308930-e3c5-47bb-bcee-58eaae7d3dfa 062_Ntr_BRCA_TCGA-BH-A1EO-11A_a426d9c2-86b1-4db1-b49c-eceaa01273a9 062_Tex_BRCA_TCGA-BH-A1EO-01A_7787e3e2-f604-4b4d-a3bc-60c795d4177b 062_Ttr_BRCA_TCGA-BH-A1EO-01A_e31bd4a4-ecda-49b4-83b0-7f1496c2f9ae	0.5849	0.48	0.8854	0.9	0.7493
63	063_Nex_BRCA_TCGA-BH-A18V-11A_353e7fa1-08c5-400a-b352-b5325e40d66c 063_Ntr_BRCA_TCGA-BH-A18V-11A_e3c5cba8-e3ba-4e0b-929b-280708e0a855 063_Tex_BRCA_TCGA-BH-A18V-01A_abcf2a8e-6f4c-4668-9ef9-41d95d16e8e6 063_Ttr_BRCA_TCGA-BH-A18V-01A_286394db-7d5e-4de2-b386-581352164350	0.6941	0.56	NaN	0.75	NaN
64	064_Nex_BRCA_TCGA-BH-A0DT-11A_6dde640a-1d79-4e7d-9491-c500b8183d9a 064_Ntr_BRCA_TCGA-BH-A0DT-11A_71aa4cd6-75ea-4e10-b16c-ea9adbf31a98 064_Tex_BRCA_TCGA-BH-A0DT-01A_9d93c6fb-336a-4cb4-9f33-8557456753b1 064_Ttr_BRCA_TCGA-BH-A0DT-01A_61ad7408-dacd-4913-a479-c456e8b03191	0.7462	0.42	NaN	0.55	0.6659
65	065_Nex_BRCA_TCGA-GI-A2C9-11A_454dbdba-da53-4e99-9670-dff1e5bbb77c 065_Ntr_BRCA_TCGA-GI-A2C9-11A_d8aa0349-d74e-4891-8398-6476eb1935f0 065_Tex_BRCA_TCGA-GI-A2C9-01A_2f2b0909-488b-4fa3-8251-2ef6e7d5869e 065_Ttr_BRCA_TCGA-GI-A2C9-01A_01ea694e-989b-4a35-9397-5e508656d1d8	0.8062	0.51	0.8386	0.8	0.6969
66	066_Nex_BRCA_TCGA-BH-A209-11A_c580c610-832d-45be-9963-06bb918ede73 066_Ntr_BRCA_TCGA-BH-A209-11A_b8b48554-ca2f-466d-85b0-9d473cca8ca7 066_Tex_BRCA_TCGA-BH-A209-01A_7b85ca36-2fe9-4156-99eb-f79463dbc572 066_Ttr_BRCA_TCGA-BH-A209-01A_b2cf947a-5ed1-4e24-8752-bf2a6eca895a	0.5979	0.24	0.6243	0.6	0.4645
67	067_Nex_BRCA_TCGA-BH-A1EV-11A_f4d30842-7873-46d3-8f25-ae7d05909175 067_Ntr_BRCA_TCGA-BH-A1EV-11A_73e296db-9eec-4060-97c9-80a98dbb9fb6 067_Tex_BRCA_TCGA-BH-A1EV-01A_fb502696-cb13-487a-a70a-6ceefcf20ca0 067_Ttr_BRCA_TCGA-BH-A1EV-01A_93ab3adf-7ab9-455e-9007-9f51443352fe	0.8435	0.61	0.9517	0.8	0.817
68	068_Nex_BRCA_TCGA-GI-A2C8-11A_836e4482-11c7-4422-a2e5-cac9b846ea71 068_Ntr_BRCA_TCGA-GI-A2C8-11A_580d9a3d-e198-4e7b-aa1f-419d868bb0b5 068_Tex_BRCA_TCGA-GI-A2C8-01A_146c0ba4-6761-446c-be7c-7e56c0ffa37b 068_Ttr_BRCA_TCGA-GI-A2C8-01A_c0ee6e25-02b9-4b2f-9f23-fd61eedf9945	0.601	0.48	0.7846	0.85	0.6998
69	069_Nex_BRCA_TCGA-BH-A0BM-11A_92faafbd-6a76-4116-80fc-a15767aa81d0 069_Ntr_BRCA_TCGA-BH-A0BM-11A_ae127be2-5e4c-4b7e-9cf8-3aa9e529baaa 069_Tex_BRCA_TCGA-BH-A0BM-01A_c513ed81-255f-43b0-b8aa-984326201745 069_Ttr_BRCA_TCGA-BH-A0BM-01A_006b2b95-7069-4cb6-bfe8-7edb80056add	0.796	0.61	0.9189	0.84	0.7886
70	070_Nex_BRCA_TCGA-BH-A18L-11A_7a010ccd-f780-45f0-98da-cc738e87b6d3 070_Ntr_BRCA_TCGA-BH-A18L-11A_ef4660c0-c177-4d46-90f9-56e3dc47b59e 070_Tex_BRCA_TCGA-BH-A18L-01A_0d4aca9c-c11e-4f78-a250-08d45ce4828e 070_Ttr_BRCA_TCGA-BH-A18L-01A_1af43803-7afa-4d2b-aa78-2dec84c1e702	0.927	0.81	NaN	0.8	0.9113
71	071_Nex_BRCA_TCGA-A7-A13F-11A_471d1e10-7c79-44f9-a373-bd3e510b6155 071_Ntr_BRCA_TCGA-A7-A13F-11A_4e4cd9e5-27bb-4ea7-9328-0b267373ec1c 071_Tex_BRCA_TCGA-A7-A13F-01A_d8fad6b2-66b8-4d6f-b018-653998675921 071_Ttr_BRCA_TCGA-A7-A13F-01A_75898a6d-75e4-4dca-a7ed-c11056e0c9c4	0.8478	0.65	0.9236	0.7	0.7915
72	072_Nex_BRCA_TCGA-E2-A15M-11A_b2138cda-519f-4691-bf1c-0f863b55d888 072_Ntr_BRCA_TCGA-E2-A15M-11A_bf873756-8ee8-49bd-b2ca-17223c7ef962 072_Tex_BRCA_TCGA-E2-A15M-01A_1ccf392d-7959-4bb9-8ca8-4298409f4951 072_Ttr_BRCA_TCGA-E2-A15M-01A_b2569032-6147-4a1c-973f-d5985127e9f4	0.4911	0.28	NaN	0.8	0.4285