CAFE MOCHA: An Integrated Platform for Discovering Clinically Relevant Molecular Changes in Cancer; an Example of Distant Metastasis and Recurrence-linked Classifiers in Head and Neck Squamous Cell Carcinoma

Neeraja M Krishnan[1], Mohanraj I[1], Janani Hariharan[1], and Binay Panda[1,2*]

[1]Ganit Labs, Bio-IT Centre, Institute of Bioinformatics and Applied Biotechnology, Bangalore, Karnataka 560100, India

[2]Strand Life Sciences, Bangalore, Karnataka 560024, India

*To whom correspondence should be addressed. Email: binay@ganitlabs.in

Abstract

Background

CAFE MOCHA (Clinical Association of Functionally Established MOlecular CHAnges) is an integrated GUI-driven computational and statistical framework to discover molecular signatures linked to a specific clinical attribute in a cancer type. We tested *CAFE MOCHA* in head and neck squamous cell carcinoma (HNSCC) for discovering a signature linked to distant metastasis and recurrence (MR) in 517 tumors from TCGA and validated the signature in 18 tumors from an independent cohort.

Methods

The platform integrates mutations and indels, gene expression, DNA methylation and copy number variations to discover a classifier first, predict an incoming tumour for the same by pulling defined class variables into a single framework that incorporates a coordinate geometry-based algorithm, called Complete Specificity Margin Based Clustering (CSMBC) with 100% specificity. *CAFE MOCHA* classifies an incoming tumour sample using either a matched normal or a built-in database of normal tissues. The application is packed and

deployed using the *install4j* multi-platform installer.

Results

We tested *CAFE MOCHA* to discover a signature for distant metastasis and recurrence in HNSCC. The signature MR44 in HNSCC yielded 80% sensitivity and 100% specificity in the discovery stage and 100% sensitivity and 100% specificity in the validation stage.

Conclusions

CAFE MOCHA is a cancer type- and clinical attribute-agnostic computational and statistical framework to discover integrated molecular signature for a specific clinical attribute.

CAFE MOCHA is available in GitHub (https://github.com/binaypanda/CAFEMOCHA).

Key Words: mutation, methylation, gene expression, copy number variation (CNV), sensitivity, specificity and integrated platform.

Introduction

Cancer progression is linked to molecular changes at multiple levels, such as somatic mutations, gene expression, DNA methylation and copy number changes. In the last 5yrs, a large amount of data on key variants in multiple cancers has been generated by international consortia, like The Cancer Genome Atlas (TCGA) (Editorial, 2015; Weinstein *et al*, 2013), International Cancer Genome Consortium (ICGC) (Hudson *et al*, 2010) and individual laboratories, aiding our understanding of various cancers at molecular level. Utilizing the vast amount of molecular data from studies on tumour genomes, exomes, transcriptomes and methylomes, and linking those with data from genetic and functional studies will help find clinically relevant insights. Although there are existing databases and studies that combine molecular changes across cancer types (Deng *et al*, 2016; Huang *et al*, 2015; Netanely *et al*, 2016), studies linking the events explicitly within the same tumour type across a large number of samples to predict signatures for a specific clinical attribute are currently lacking.

Here, we describe a cancer type-agnostic computational and statistical framework with a user-friendly graphical-user-interface (GUI) to discover integrated signatures using six tumour-specific event types; somatic mutations and indels (mut), DNA copy number changes (cnv), gene expression changes (expr), 5-cytosine DNA methylation changes (meth), functional copy number changes (fcnv, where CNVs are linked to gene expression changes) and functional *cis*-regulatory DNA methylation changes (fmeth, where hyper- and hypo-methylation result in down- and up-regulation of effector gene expression respectively). *CAFE MOCHA*, in addition to classifying categorical events like mut and cnv, uses an algorithm called Complete Specificity Margin Based Clustering (CSMBC) to classify quantitative (expr and meth) and coupled events (fmeth and fcnv), and combines the event types using sample frequency and event priority filters, to produce integrated

signatures describing the clinical variable with 100% specificity. We tested *CAFE MOCHA* to discover a signature linked with distant metastasis and recurrence in head and neck squamous cell carcinoma (HNSCC) using 434 tumours from TCGA as a discovery cohort, which was validated in an independent cohort of 18 tumours. The integrated signature MR44 for metastatic and recurrent tumours in HNSCC (MR44) yielded a sensitivity of 79.52% and a specificity of 100% in the discovery cohort and with 100% sensitivity and 100% specificity in the validation cohort.

Materials and Methods

Discovery module

CAFE MOCHA application workflow and the GUI are illustrated in Figure 1. *CAFE MOCHA* has two independent modules, discovery and prediction. In the discovery module, both somatic mutations (missense, nonsense and splice-site) and indels were considered under mut. A matrix of following values (thresholds) were considered for cnvs; -2: full copy deletion, -1: allelic deletion, 1: low-copy amplification, 2: high-copy amplification, and 0: lack of any copy number variation. For expression and methylation, RSEM (Li & Dewey, 2011) -normalized gene-wise intensity matrix and a probe-wise matrix of pre-processed β values (from Illumina 450k whole-genome methylation arrays), respectively, were used. The methylation data was pre-processed, including normalization and batch correction, as described previously (Krishnan *et al*, 2016). Tumour samples assayed for all the four events (mut, cnv, expr and meth) were considered for the discovery phase. The perturbed genes were passed through a functional filter (detailed below) before being used in the integrated analysis. In the case of methylation data, a sub-region filter was applied and only probes located in the gene promoters, transcription start site (TSS200 and TSS1500) and CpG islands were retained. Categorical events like mut and cnvs were linked to the

chosen clinical attributes with complete specificity. Quantitative events such as expression (expr) and methylation (meth) were linked to clinical attributes using an algorithm called Complete Specificity Margin Based Clustering (CSMBC) algorithm (detailed below). A somatic filter based on an unpaired *t*-test ($P < 0.05$) was further used to retain only tumour-specific expr and meth events. Expression events were further coupled with meth and cnvs to generate coupled events (fmeth and fcnv) respectively, where meth and cnvs affected the target gene expression. We only considered hyper- and hypo-methylation events linked with the down- and up-regulation of downstream gene expression respectively.

Data integration and making of the final signature

Using the 'integrate' feature, the individually discovered events were combined into an integrated signature with two different filters (Figure 1B). In the first, a tumour-specific sample frequency-based filter agnostic to any event type was applied. In the second, all event types' priorities were determined by detection sensitivity first and then those events that yield highest priority followed by highest sample frequency were chosen. These two filters work contrarily, one retaining the high sample frequency quantitative and semi-quantitative event types (expr, meth, fmeth, fcnv) and another the lower-sample frequency categorical and semi-categorical event types (mut, cnv, fcnv). Selected events were considered for removal of false positives using samples lacking complete cross-platform overlap from TCGA and the independently provided datasets (used for confirmation in the second stage of the discovery stage). The number of confirmed events was further minimized using the sample frequency-based filter to constitute the discovery panel for that specific clinical attribute. This minimal signature is then subjected to an independent validation using tumours assayed for all the four events (mut, cnv, expr, meth) (Figure 1B). Finally, the discovered events were mapped to all of the sixteen cancer-related pathways

121 from KEGG (Kanehisa *et al*, 2002; Kanehisa *et al*, 2004; Kanehisa *et al*, 2012)
122 (hsa05200).
123
124
125 *CAFE MOCHA Interface*
126 The interface for *CAFE MOCHA* was developed and deployed using *Netbeans IDE v*
127 *8.1* (https://netbeans.org/downloads/). The Graphical User Interface (GUI) was designed
128 using JAVA AWT (http://www.javatpoint.com/java-awt) with SWING components
129 (http://www.javatpoint.com/java-swing) with its native OS GUI and appropriate controls.
130 The back end for this interface was built on a Linux-dependent platform with *R*, *BASH*,
131 *BEDTools* (v2.3) and *PERL* as added dependencies. *CAFE MOCHA* requires installations
132 of *R* package dependencies (minfi v1.18.2 (Aryee *et al*, 2014), wateRmelon v1.16.0
133 (Pidsley *et al*, 2013), IlluminaHumanMethylation450kmanifest v0.4.0, randomForest
134 v4.6.12(Breiman, 2001), varSelRF v0.7.5, pheatmap v1.0.8, gplots v3.0.1, ggplot2 v2.1.0 ,
135 reshape2 v1.4.1) to pre-process 450k *idat* files and generates the matrix of β values. The
136 platform provides the user with an option to install the dependencies. Under the discovery
137 module, mut, cnv, expr and meth data can be browsed and acquired locally.
138
139 *Functional filters*
140 To ensure functional relevance of predicted tumour-specific molecular changes, we
141 introduced various filters for different events to select genes of interest. We used IntOGen
142 (Gundem *et al*, 2010) to select driver/potential driver mutations, MethHC (Huang *et al*,
143 2015) for known hyper- and hypomethylated genes in different cancer types, G2SBC
144 (Mosca *et al*, 2010) for selecting important differentially expressed genes associated with
145 breast cancer, The Human Protein Atlas (Uhlen *et al*, 2015) for functional proteins, the
146 dbDEPC 2.0 database (He *et al*, 2012) for manually curated differentially expressed

147 proteins in human cancer and miRTarBase (Chou *et al*, 2016; Hsu *et al*, 2011; Hsu *et al*,
148 2014) for experimentally validated miRNA-target gene associations. In the case of micro-
149 RNA methylation, we considered those events where methylation of a microRNA
150 correlated with a concomitant expression change in one or more of its target gene(s). In
151 the case of fcnv, we considered where amplification and deletion in a gene was linked to
152 its up- and down-regulation respectively in tumour sample compared to normal. Only full
153 copy deletions (not allelic deletions) and amplifications were considered. Genes bearing
154 categorical events in at least one sample that passed the functional filter were considered
155 further.
156
157 *Complete Specificity Margin Based Clustering (CSMBC)*
158 We devised a statistically framework, Complete Specificity Margin Based Clustering
159 or CSMBC, for inferring the expression and the methylation boundaries of each cluster for
160 a specific clinical attribute (Figure 2). In CSMBC, the number of clusters were fixed
161 equivalent to the number of pre-defined clinical variables. For each cluster, axis-
162 perpendicular boundaries passing through the most extreme outliers on the
163 expression/methylation coordinate axes were determined first. The number of overlaps
164 was dependent upon the number of boundaries. In the simplest case, there can be two
165 sub-categories. For example, metastasis as a category has two sub-categories, metastatic
166 (A); and non-metastatic (B). In this simplest case, there can only be one overlap ($A \cap B$).
167 Likewise, if a clinical attribute has three (A, B and C) sub-categories (for example
168 recurrent, non-recurrent, can't be defined) with four possible overlaps
169 ($A \cap B, A \cap C, B \cap C, A \cap B \cap C$). Using the above logic defined for categories and sub-
170 categories, all possible overlaps between clusters were identified first, and then the
171 minimum and maximum values for various overlaps were estimated according to the
172 following equations:

For overlap between two sub-categories:

$$i \cap_{i \neq j} j_{min} = max(i, j); i \cap_{i \neq j} j_{max} = min(i, j)$$

For overlap between three sub-categories:

$$i \cap_{i \neq j \neq k} j \cap k_{min} = max(i, j, k); i \cap_{i \neq j \neq k} j \cap k_{max} = min(i, j, k)$$

For overlap between four sub-categories:

$$i \cap_{i \neq j \neq k \neq l} j \cap k \cap l_{min} = max(i, j, k, l); i \cap_{i \neq j \neq k \neq l} j \cap k \cap l_{max} = min(i, j, k, l)$$

A sample was identified to be within 100% specificity margin of the cluster if it was not in any region of overlap with other clusters (Figure 2). The total number of all such samples, *i.e.,* sensitivity of that cluster, was estimated according to the equation below:

For two sub-categories:

$$sensitivity_i = n_i - n_{i \cap j}$$

For three sub-categories:

$$sensitivity_i = n_i - n_{i \cap j} - n_{i \cap k} + n_{i \cap j \cap k}$$

For four sub-categories:

$$sensitivity_i = n_i - n_{i \cap j} - n_{i \cap k} - n_{i \cap l} + n_{i \cap j \cap k} + n_{i \cap j \cap l} + n_{i \cap k \cap l} - n_{i \cap j \cap k \cap l}$$

Finally, an unpaired *t*-test was performed between all tumour and all normal samples, and genes/probes with significantly different expr/meth values ($P < 0.05$) between the tumour and normal groups were retained.

Prediction module

The prediction module has two options: one, where the user provides somatic mutations/indels, tumour-specific expr, meth and cnv information, and second, where the user has somatic mutations/indel/cnv data but does not have expr and meth data from the matched normal samples. In the later case, *CAFE MOCHA* uses a built in normal database (all normal samples from the TCGA for a particular cancer type) to compute expression and methylation values for genes in control samples from RNASeq for expression and whole genome arrays (Illumina 450k) for DNA methylation. For the prediction module, the data is entered as per the pre-defined format (same as discovery panel) specific to a cancer type and clinical attribute of interest (Figure 1). Based on the user input, coupled events (fmeth and fcnv) are determined and used for the prediction module.

Testing CAFE MOCHA to discover metastatic/recurrent HNSCC signature

We tested *CAFE MOCHA* to discover integrated signature for metastatic and recurrent HNSCC tumours using tumours from TCGA dataset (n=434), followed by confirmation (removal of false positives, the second stage of the discovery phase) using two sample cohorts, 42 samples from TCGA and 37 samples from an independent cohort (Krishnan *et al*, 2015; Krishnan *et al*, 2016), where the information on at least one of the event type out of the 4 events (mut, cnv, expr and meth) for the same tumour was available (Table 1, Supplementary Tables S1 and S2). Finally, the discovery panel was validated in 18 samples from an independent sample cohort (Krishnan *et al*, 2015; Krishnan *et al*, 2016), where all the four events were assayed for all the tumours (Supplementary Table S3). For the discovery module, we downloaded the Broad automated somatic mutations and indel call file, copy number data with gistic2 threshold, Illumina HiSeqV2 RSEM (7) -normalized RNA-seq gene expression data, 450k DNA methylation data and the clinical data from the

UCSC Xena Browser (http://tinyurl.com/jhmg9b9). The distant metastasis (M1) and recurrent tumours (RFS_IND = 1) were compared with non-metastatic (M0) and non-recurrent (RFS_IND = 0) tumours to derive a specific signature.

Results

Discovery of distant metastasis/recurrence-associated molecular signature MR44 in HNSCC

CAFE MOCHA identified a total of 6649 events from individual event type discovery (8 somatic mutations, 1436 cnvs (506 amplifications and 982 full copy deletions), 1098 expression-related events, 4018 methylation events, 27 fcnv events (8 high copy amplifications and 19 full copy deletions) and 62 fmeth events), associated with metastasis and recurrence in HNSCC. Integration across discovered events was performed using two filters as described in the Methods section. The filters resulted in 79 (22 expression-related events and 57 methylated genes) and 171 (7 mut, 84 cnvs (46 amplifications and 38 full copy deletions), 24 expr, 2 meth, 25 fcnv events (8 high copy amplifications and 17 full copy deletions) and 29 fmeth events) events, individually, and 232 (7 mut, 84 cnvs (46 amplifications and 38 full copy deletions), 30 expr, 57 meth, 25 fcnv events (8 high copy amplifications and 17 full copy deletions) and 29 fmeth events), cumulatively after removing any redundancy (Supplementary Table S4).

Using data at the second stage of the discovery phase (confirmation stage), 98 false positives events were eliminated. The remaining 134 events (Supplementary Figure S1) (5 mut, 83 cnvs (46 amplifications and 37 full copy deletions), 8 expr, 9 meth, 24 fcnv events (8 high copy amplifications and 16 full copy deletions) and 5 fmeth events) were further minimized to a final discovery panel of 44 events (MR44) that included 2 mut, 15 cnvs (3 amplifications and 12 full copy deletions), 8 expr, 9 meth, 8 fcnv events (5 high

copy amplifications and 3 full copy deletions) and 2 fmeth events) (Figure 3A). The signature was further validated using an independent cohort of 18 samples (3 M1/R and 15 non-MR) of an independent oral tongue squamous cell carcinoma (OTSCC) cohort, with 100% sensitivity and 100% specificity. 17 out of the 44 events in MR44 signature were validated in at least one sample in the validation cohort (Figure 3B).

Power of integration

We wanted to investigate the power of integration (where all the six different event types were used to derive the signature) over lesser order combinations or with individual event type, in terms of the detection sensitivity and the number of events required to achieve maximum specificity. As shown in Figure 4, the detection sensitivity showed a gradual enhancement when the number of event types increased from one to six. For single-event analyses, meth and both mut and fmeth provided with a maximum and minimum sensitivity respectively (37%, 7%, Figure 4A). The number of CNVs required to achieve the highest possible level of sensitivity in a single-event analysis was the highest (83 events with 33% sensitivity). When using more than one event type, we obtained an increase in sensitivity to various levels gradually from a single-event to a six-events signature (Figure 4). The number of CNVs and both mut and fmeth required to get similar level of sensitivity was highest and lowest respectively for all combinations. Highest sensitivity (80%) of detection was observed when all the six different event types were used to produce the integrated signature (Figure 4F). The impact of event integration on detection sensitivity was observed to be greatest when the integration was performed in the following order: meth + cnvs + expr + fmeth + fcnvs + mut. Each incremental integration in this order caused a gradual increase in sensitivity (37.35 → 56.63 → 68.67 → 74.70 → 77.11 → 79.52, Figure 4). The actual events for all combinations and individual event type analyses are illustrated in Supplementary Figure S1.

Discussion

Integrating multiple molecular events is a strategy proposed to identify master regulators in cancer (Gevaert & Plevritis, 2013; Thingholm *et al*, 2016). However, a meaningful integration should be incremental, with each level filtering out the unnecessary and only retaining the functionally relevant associations. Furthermore, integration of data should link a specific signature to clinically relevant tumour attributes in order for making meaningful conclusions. Integrated analyses often fail to establish the link between cancer-associated changes within the same tumour types across cohorts, perhaps due to the low overlapping of clinical characteristics across cohorts and low reproducibility of results across laboratories, geographies and discovery platforms. Additionally, data on multiple events; genetic, transcriptomic and epigenetic changes, from the same sets of tumour:matched normal samples are often not available , making it difficult to discover truly integrated signatures. Nevertheless, such a discovery set, where all the events are assayed for all tumour:normal pairs within the same cohort, is functionally more meaningful. As multidimensionality increases with the availability of more data, especially from the large consortia like TCGA and ICGC, it will become imperative to design and implement easy-to-use and robust web-based platforms to discover (using a pre-defined training set) multi-gene and multi-platform classifiers for a particular clinical attribute and predict the same for an incoming new tumour sample. Currently, such a platform is lacking that can discover tumour-specific genome-wide functional and somatic molecular changes linked to a specific clinical attribute, using a sizeable multi-dimensional dataset. Keeping this in mind, we devised *CAFE MOCHA*, an automated and integrated framework to discover meaningful, functional and somatic molecular changes in a cancer type that links the signature(s) to a specific clinical attribute. *CAFE MOCHA* is designed to use both user-defined/generated and publicly available tumour and matched normal data.

Additionally, the prediction module of *CAFE MOCHA* uses a backend database of unmatched normal samples to increase the prediction scope of the tool for tumours where matched normal samples are not available, especially where expression and methylation data is not available from matched normal samples. *CAFE MOCHA* is a cancer type- and clinical attribute-agnostic tool and has the ability to pull data on several event types under a single framework. *CAFE MOCHA* classifies events into three different types, categorical (mutations and CNV), quantitative (expression and methylation) and coupled/linked (functional CNVs and functional methylation) and uses an algorithm called complete specificity margin based clustering (CSMBC), a fully supervised approach, to identify clinically linked quantitative and coupled events. CSMBC is a modification of the Large Rectangle Margin Learning (LRML) approach described previously (kirmse & Petersohn, 2011) and is conceptually similar to fast boxes which take the 'characterize' and then 'discriminate' approach of classification (Goh & Rubin, 2014). The input for CSMBC is a continuous function spread across a single dimension, either expression or methylation. For both these variables, an unsupervised clustering, which requires *a priori* knowledge of the quantitative expression or methylation map, is not feasible since these are continuous variables. The selection of boundary margins, according to the chosen clinical variables on the other hand does not require any prior knowledge of the quantitative spread of these events. Additionally, supervised clustering is faster than unsupervised clustering since it does not require iterative computing and does not require re-estimation of the cluster mean. Nevertheless, as pointed out in the past (Richards, 2013), a hybrid semi-supervised approach that uses the results of an unsupervised approach such as the *k*-means or Expectation Maximization (EM) as training areas for a supervised classification, might combine the advantages of both supervised and unsupervised classification approaches. A weighted sensitivity method (Gao, 2007; Iwamoto & Pusztai, 2010) that weighs the

328 classification sensitivity of a sample based on its proximity to the nearest boundary, would
329 further add confidence perhaps at the cost of sensitivity, particularly while predicting the
330 clinical status of an incoming sample. While we plan to develop *CAFE MOCHA* taking
331 these modifications into account in the future, the current implementation benefits from
332 ensuring complete specificity for each clinical attribute-associated molecular event. One of
333 the limitations of CSMBC is that its specificity is entirely dependent on the denominator,
334 *i.e.,* sample numbers and therefore, may be sub-optimal in a scenario where fewer
335 numbers of tumour samples are present in the discovery set.
336
337 HNSCC are the sixth leading cause of cancer worldwide with 5-year survival of less
338 than 50% (Ferlay *et al*, 2010; Mishra & Meherotra, 2014). Recent high-throughput studies
339 employing computational methods have identified various genetic, transcriptomic and
340 epigenetic changes from different subsites of HNSCC from different geographies (Agrawal
341 *et al*, 2011; Cancer Genome Atlas, 2015; India Project Team of the International Cancer
342 Genome, 2013; Krishnan *et al*, 2015; Pickering *et al*, 2013). Early-stage patients with
343 HNSCC are usually treated using a single modality like surgery or radiotherapy whereas
344 advanced-stage patients benefit from multi-modality therapies (Ridge *et al*, 2016). In head
345 and neck tumours, identifying patients with tumours *a priori* that may undergo distant
346 metastasis or loco-regional recurrence using primary-tumour-derived molecular signature
347 will help manage patients better. *CAFE MOCHA* was tested using a robust dataset of 434
348 HNSCC tumours from TCGA where data on all four molecular events were available.
349 Despite the MR44 being an integrated signature, drawn from 83 metastatic/recurrent
350 tumours with four different somatic events (mut, cnv, expr and meth) and two coupled
351 events (fmeth and fcnv) available on all the tumours, the discovery sensitivity did not attain
352 close to 100% (sensitivity was 79.52%). This means that ~20% of the metastatic/recurrent
353 tumours could not be classified using the integrated MR44 signature. This could indicate

two possibilities: first, there are other types of tumour-specific changes that need to be assayed and accounted for in order to derive a truly complete distant metastasis and recurrence-associated signature, and/or, second: the sample size of 83 metastatic/recurrent tumours was not sufficient (the predictive power of discovery was not optimum). The fact that the validation sensitivity of MR44 was 100% in an independent cohort does not deter from either of these conclusions as the number of tumours in the validation set was small (n = 18).

About half of the genes in the MR44 signature have been reported previously to be associated with prognosis, survival, recurrence and distant metastasis for various cancer types. For example, *ITGA9* is reported as a tumour suppressor gene in non-small cell lung cancer (Pastuszak-Lewandoska *et al*, 2016) and is involved in cell migration and invasion in melanoma (Zhang *et al*, 2016). Additionally, epigenetic inactivation of *ITGA9* is linked with its expression in nasopharyngeal and breast cancer (Mostovich *et al*, 2011; Nawaz *et al*, 2015). *NRAS* is linked with survival of patients with liver metastases in colorectal cancer (Vauthey *et al*, 2013) and acts as a prognostic factor in metastatic melanoma (Jakob *et al*, 2012). The chromosomal region containing *DVL2* gene frequently undergoes LOH in patients with colorectal tumours (Kurashina *et al*, 2008). *RPL26A1* expression was a prognostic marker in ER-positive breast tumours (Wang & Zhang, 2007) and liver metastases in colorectal cancer (Nakamura & Furukawa, 2003). *MORF4L1* was found to be one of the candidate genes with copy number reductions in breast cancer (Chen *et al*, 2007). *FGFR4* profile was observed to be prognostically relevant in squamous cell carcinoma of the mouth and oropharynx (Dutra *et al*, 2012), esophagus (Shim *et al*, 2016), and gastric cancer (Murase *et al*, 2014). *ANXA11* expression is one of the prognostic markers in prostate cancer (Chandran *et al*, 2007; Tsai *et al*, 2013). *ZBTB7A* plays a role in suppressing tumour metastasis in gastric cancer (Shi *et al*, 2015). Loss of stromal *CAV1*

expression predicts poor survival in colorectal cancer (Zhao *et al*, 2015). *ARHGEF38* expression significantly differed between high and low recurrence-free survival groups in prostrate cancer (Tandefelt *et al*, 2013). *OXCT1* was shown to act as an oncogene in human breast cancer cells (Martinez-Outschoorn *et al*, 2012). *RICTOR* regulates cancer cell metastases in breast cancer cell lines (Zhang *et al*, 2010). *RPAP2* associates with breast cancer recurrence (Baker *et al*, 2014). mIR573 inhibits prostate cancer metastases (Wang *et al*, 2015). *AIMP1* is down regulated in gastric and colorectal cancer (Kim *et al*, 2014). *FAM134C* is a cancer-relapse marker in breast and ovarian cancer (Guo, 2011). *LIMS1* (PINCH signalling) has been linked to distant metastasis in pancreatic stromal cells (Scaife *et al*, 2010). *EGLN1* (commonly referred to as *PHD2*) is an oxygen sensor that promotes breast cancer metastases (Kuchnio *et al*, 2015). *RPLP2* is one of the 10-gene prognostic markers for gastric cancer (Zhang *et al*, 2011). *TOLLIP* was found to be significantly and differentially expressed in colorectal metastatic cells (Barderas *et al*, 2013). *ZNF490* was found to be associated with breast cancer recurrence (Baker *et al*, 2013). MR44 contains genes implicated in MTOR, MAPK and PI3K-AKT signalling pathways via *DVL2*, *FGF10*, *FGFR4*, *RICTOR*, *ITGA9*, *NRAS* and *CAV1* genes (Supplementary Table S4).

Although the discovery module of *CAFE MOCHA* uses 450k methylation array and RSEM-normalized RNASeq expression data, the predict module is not restricted to the assays being performed on the same platform so long as the user provides data in the same format for a matched normal. Future versions of *CAFE MOCHA* will incorporate cross-platform converter wherein the user can deposit data generated with multiple platforms and in multiple formats, thus making the tool truly platform agnostic. Additionally, *CAFE MOCHA* has the ability to use an underlying tissue-specific normal database, especially where expression and methylation data is not available, for un-matched normal samples. However, if data for a matched normal is not available, the user must use the

same assays and formats, used in the discovery module to minimize assay/platform related errors.

Funding

Research presented in this article is funded by Department of Electronics and Information Technology, Government of India (Ref No:18(4)/2010-E-Infra., 31-03-2010) and Department of IT, BT and ST, Government of Karnataka, India (Ref No:3451-00-090-2-22).

References

Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, Fakhry C, Xie TX, Zhang J, Wang J, Zhang N, El-Naggar AK, Jasser SA, Weinstein JN, Trevino L, Drummond JA, Muzny DM, Wu Y, Wood LD, Hruban RH, Westra WH, Koch WM, Califano JA, Gibbs RA, Sidransky D, Vogelstein B, Velculescu VE, Papadopoulos N, Wheeler DA, Kinzler KW, Myers JN (2011) Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* **333**(6046)**:** 1154-7

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA (2014) Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**(10)**:** 1363-9

Baker JB, Sinicropi DV, Pelham RJ, Crager M, Collin F, Stephans JC, Liu M, Morlan J, Qu K (2013) Method of predicting breast cancer prognosis.

Baker JB, Sinicropi DV, Pelham RJ, Crager M, Collin F, Stephans JC, Liu M, Morlan J, Qu K (2014) Method of predicting breast cancer prognosis

Barderas R, Mendes M, Torres S, Bartolome RA, Lopez-Lucendo M, Villar-Vazquez R, Pelaez-Garcia A, Fuente E, Bonilla F, Casal JI (2013) In-depth characterization of the secretome of colorectal cancer metastatic cells identifies key proteins in cell adhesion, migration, and invasion. *Mol Cell Proteomics* **12**(6)**:** 1602-20

Breiman L (2001) *Random Forests*. Vol. 45. The Netherlands: Kliwer Academic Publishers

Cancer Genome Atlas N (2015) Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**(7536)**:** 576-82

Chandran UR, Ma C, Dhir R, Bisceglia M, Lyons-Weiler M, Liang W, Michalopoulos G, Becich M, Monzon FA (2007) Gene expression profiles of prostate cancer reveal involvement of multiple molecular pathways in the metastatic process. *BMC Cancer* **7:** 64

Chen W, Salto-Tellez M, Palanisamy N, Ganesan K, Hou Q, Tan LK, Sii LH, Ito K, Tan B, Wu J (2007) Targets of genome copy number reduction in primary breast cancers identified by integrative genomics. *Genes, Chromosomes and Cancer* **46**(3)**:** 288-301

Chou CH, Chang NW, Shrestha S, Hsu SD, Lin YL, Lee WH, Yang CD, Hong HC, Wei TY, Tu SJ, Tsai TR, Ho SY, Jian TY, Wu HY, Chen PR, Lin NC, Huang HT, Yang TL, Pai CY, Tai CS, Chen WL, Huang CY, Liu CC, Weng SL, Liao KW, Hsu WL, Huang HD (2016) miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic acids research* **44**(D1)**:** D239-47

Deng M, Bragelmann J, Schultze JL, Perner S (2016) Web-TCGA: an online platform for integrated analysis of molecular cancer data sets. *BMC bioinformatics* **17:** 72

Dutra RL, de Carvalho MB, Dos Santos M, Mercante AM, Gazito D, de Cicco R, Group G, Tajara EH, Louro ID, da Silva AM (2012) FGFR4 profile as a prognostic marker in squamous cell carcinoma of the mouth and oropharynx. *PLoS One* **7**(11)**:** e50747

Editorial (2015) The future of cancer genomics. *Nature medicine* **21**(2)**:** 99

Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International journal of cancer Journal international du cancer* **127**(12)**:** 2893-917

Gao BJ. Hyper-rectangle-based discriminative data generalization and applications in data mining. Ph.D., Hyper-rectangle-based discriminative data generalization and applications in data mining. Ph.D. Thesis. Simon Fraser University, Canada, 2007

Gevaert O, Plevritis S (2013) Identifying master regulators of cancer and their downstream targets by integrating genomic and epigenomic features. *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing***:** 123-34

Goh ST, Rubin C (2014) Box Drawings for Learning with Imbalanced Data *aRxiv [statML]*

Gundem G, Perez-Llamas C, Jene-Sanz A, Kedzierska A, Islam A, Deu-Pons J, Furney SJ, Lopez-Bigas N (2010) IntOGen: integration and data mining of multidimensional oncogenomic data. *Nature methods* **7**(2)**:** 92-3

Guo NL (2011) Gene signature for diagnosis and prognosis of breast cancer and ovarian cancer: Google Patents. US8030060

He Y, Zhang M, Ju Y, Yu Z, Lv D, Sun H, Yuan W, He F, Zhang J, Li H, Li J, Wang-Sattler R, Li Y, Zhang G, Xie L (2012) dbDEPC 2.0: updated database of differentially expressed proteins in human cancers. *Nucleic acids research* **40**(Database issue)**:** D964-71

Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, Chan WL, Tsai WT, Chen GZ, Lee CJ, Chiu CM, Chien CH, Wu MC, Huang CY, Tsou AP, Huang HD (2011) miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic acids research* **39**(Database issue)**:** D163-9

Hsu SD, Tseng YT, Shrestha S, Lin YL, Khaleel A, Chou CH, Chu CF, Huang HY, Lin CM, Ho SY, Jian TY, Lin FM, Chang TH, Weng SL, Liao KW, Liao IE, Liu CC, Huang HD (2014) miRTarBase update 2014: an information resource for experimentally validated

502 miRNA-target interactions. *Nucleic acids research* **42**(Database issue)**:** D78-85
503
504 Huang WY, Hsu SD, Huang HY, Sun YM, Chou CH, Weng SL, Huang HD (2015) MethHC:
505 a database of DNA methylation and gene expression in human cancer. *Nucleic acids*
506 *research* **43**(Database issue)**:** D856-61
507
508 Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabe RR, Bhan MK, Calvo F,
509 Eerola I, Gerhard DS, Guttmacher A, Guyer M, Hemsley FM, Jennings JL, Kerr D, Klatt P,
510 Kolar P, Kusada J, Lane DP, Laplace F, Youyong L, Nettekoven G, Ozenberger B,
511 Peterson J, Rao TS, Remacle J, Schafer AJ, Shibata T, Stratton MR, Vockley JG,
512 Watanabe K, Yang H, Yuen MM, Knoppers BM, Bobrow M, Cambon-Thomsen A, Dressler
513 LG, Dyke SO, Joly Y, Kato K, Kennedy KL, Nicolas P, Parker MJ, Rial-Sebbag E, Romeo-
514 Casabona CM, Shaw KM, Wallace S, Wiesner GL, Zeps N, Lichter P, Biankin AV,
515 Chabannon C, Chin L, Clement B, de Alava E, Degos F, Ferguson ML, Geary P, Hayes
516 DN, Hudson TJ, Johns AL, Kasprzyk A, Nakagawa H, Penny R, Piris MA, Sarin R, Scarpa
517 A, Shibata T, van de Vijver M, Futreal PA, Aburatani H, Bayes M, Botwell DD, Campbell
518 PJ, Estivill X, Gerhard DS, Grimmond SM, Gut I, Hirst M, Lopez-Otin C, Majumder P,
519 Marra M, McPherson JD, Nakagawa H, Ning Z, Puente XS, Ruan Y, Shibata T, Stratton
520 MR, Stunnenberg HG, Swerdlow H, Velculescu VE, Wilson RK, Xue HH, Yang L,
521 Spellman PT, Bader GD, Boutros PC, Campbell PJ, Flicek P, Getz G, Guigo R, Guo G,
522 Haussler D, Heath S, Hubbard TJ, Jiang T, Jones SM, Li Q, Lopez-Bigas N, Luo R,
523 Muthuswamy L, Ouellette BF, Pearson JV, Puente XS, Quesada V, Raphael BJ, Sander
524 C, Shibata T, Speed TP, Stein LD, Stuart JM, Teague JW, Totoki Y, Tsunoda T, Valencia
525 A, Wheeler DA, Wu H, Zhao S, Zhou G, Stein LD, Guigo R, Hubbard TJ, Joly Y, Jones
526 SM, Kasprzyk A, Lathrop M, Lopez-Bigas N, Ouellette BF, Spellman PT, Teague JW,
527 Thomas G, Valencia A, Yoshida T, Kennedy KL, Axton M, Dyke SO, Futreal PA, Gerhard
528 DS, Gunter C, Guyer M, Hudson TJ, McPherson JD, Miller LJ, Ozenberger B, Shaw KM,
529 Kasprzyk A, Stein LD, Zhang J, Haider SA, Wang J, Yung CK, Cros A, Liang Y,
530 Gnaneshan S, Guberman J, Hsu J, Bobrow M, Chalmers DR, Hasel KW, Joly Y, Kaan TS,
531 Kennedy KL, Knoppers BM, Lowrance WW, Masui T, Nicolas P, Rial-Sebbag E, Rodriguez
532 LL, Vergely C, Yoshida T, Grimmond SM, Biankin AV, Bowtell DD, Cloonan N, deFazio A,
533 Eshleman JR, Etemadmoghadam D, Gardiner BB, Kench JG, Scarpa A, Sutherland RL,
534 Tempero MA, Waddell NJ, Wilson PJ, McPherson JD, Gallinger S, Tsao MS, Shaw PA,
535 Petersen GM, Mukhopadhyay D, Chin L, DePinho RA, Thayer S, Muthuswamy L, Shazand
536 K, Beck T, Sam M, Timms L, Ballin V, Lu Y, Ji J, Zhang X, Chen F, Hu X, Zhou G, Yang Q,
537 Tian G, Zhang L, Xing X, Li X, Zhu Z, Yu Y, Yu J, Yang H, Lathrop M, Tost J, Brennan P,
538 Holcatova I, Zaridze D, Brazma A, Egevard L, Prokhortchouk E, Banks RE, Uhlen M,
539 Cambon-Thomsen A, Viksna J, Ponten F, Skryabin K, Stratton MR, Futreal PA, Birney E,
540 Borg A, Borresen-Dale AL, Caldas C, Foekens JA, Martin S, Reis-Filho JS, Richardson
541 AL, Sotiriou C, Stunnenberg HG, Thoms G, van de Vijver M, van't Veer L, Calvo F,
542 Birnbaum D, Blanche H, Boucher P, Boyault S, Chabannon C, Gut I, Masson-Jacquemier
543 JD, Lathrop M, Pauporte I, Pivot X, Vincent-Salomon A, Tabone E, Theillet C, Thomas G,
544 Tost J, Treilleux I, Calvo F, Bioulac-Sage P, Clement B, Decaens T, Degos F, Franco D,
545 Gut I, Gut M, Heath S, Lathrop M, Samuel D, Thomas G, Zucman-Rossi J, Lichter P, Eils
546 R, Brors B, Korbel JO, Korshunov A, Landgraf P, Lehrach H, Pfister S, Radlwimmer B,
547 Reifenberger G, Taylor MD, von Kalle C, Majumder PP, Sarin R, Rao TS, Bhan MK,
548 Scarpa A, Pederzoli P, Lawlor RA, Delledonne M, Bardelli A, Biankin AV, Grimmond SM,
549 Gress T, Klimstra D, Zamboni G, Shibata T, Nakamura Y, Nakagawa H, Kusada J,
550 Tsunoda T, Miyano S, Aburatani H, Kato K, Fujimoto A, Yoshida T, Campo E, Lopez-Otin
551 C, Estivill X, Guigo R, de Sanjose S, Piris MA, Montserrat E, Gonzalez-Diaz M, Puente
552 XS, Jares P, Valencia A, Himmelbauer H, Quesada V, Bea S, Stratton MR, Futreal PA,
553 Campbell PJ, Vincent-Salomon A, Richardson AL, Reis-Filho JS, van de Vijver M, Thomas

G, Masson-Jacquemier JD, Aparicio S, Borg A, Borresen-Dale AL, Caldas C, Foekens JA, Stunnenberg HG, van't Veer L, Easton DF, Spellman PT, Martin S, Barker AD, Chin L, Collins FS, Compton CC, Ferguson ML, Gerhard DS, Getz G, Gunter C, Guttmacher A, Guyer M, Hayes DN, Lander ES, Ozenberger B, Penny R, Peterson J, Sander C, Shaw KM, Speed TP, Spellman PT, Vockley JG, Wheeler DA, Wilson RK, Hudson TJ, Chin L, Knoppers BM, Lander ES, Lichter P, Stein LD, Stratton MR, Anderson W, Barker AD, Bell C, Bobrow M, Burke W, Collins FS, Compton CC, DePinho RA, Easton DF, Futreal PA, Gerhard DS, Green AR, Guyer M, Hamilton SR, Hubbard TJ, Kallioniemi OP, Kennedy KL, Ley TJ, Liu ET, Lu Y, Majumder P, Marra M, Ozenberger B, Peterson J, Schafer AJ, Spellman PT, Stunnenberg HG, Wainwright BJ, Wilson RK, Yang H (2010) International network of cancer genome projects. *Nature* **464**(7291)**:** 993-8

India Project Team of the International Cancer Genome C (2013) Mutational landscape of gingivo-buccal oral squamous cell carcinoma reveals new recurrently-mutated genes and molecular subgroups. *Nature communications* **4:** 2873

Iwamoto T, Pusztai L (2010) Predicting prognosis of breast cancer with gene signatures: are we lost in a sea of data? *Genome medicine* **2**(11)**:** 81

Jakob JA, Bassett RL, Jr., Ng CS, Curry JL, Joseph RW, Alvarado GC, Rohlfs ML, Richard J, Gershenwald JE, Kim KB, Lazar AJ, Hwu P, Davies MA (2012) NRAS mutation status is an independent prognostic factor in metastatic melanoma. *Cancer* **118**(16)**:** 4014-23

Kanehisa M, Goto S, Kawashima S, Nakaya A (2002) The KEGG databases at GenomeNet. *Nucleic acids research* **30**(1)**:** 42-6

Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. *Nucleic acids research* **32**(Database issue)**:** D277-80

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research* **40**(1)**:** D109-14

Kim MS, Kim S, Myung H (2014) Degradation of AIMP1/p43 induced by hepatitis C virus E2 leads to upregulation of TGF-beta signaling and increase in surface expression of gp96. *PLoS One* **9**(5)**:** e96302

kirmse M, Petersohn U (2011) Large margin rectangle learning an alternative way to learn interpretable and representative models. In *International Conference of Soft Computing and Pattern Recognition (SoCPaR)* pp 161-166: IEEE

Krishnan N, Gupta S, Palve V, Varghese L, Pattnaik S, Jain P, Khyriem C, Hariharan A, Dhas K, Nair J, Pareek M, Prasad V, Siddappa G, Suresh A, Kekatpure V, Kuriakose M, Panda B (2015) Integrated analysis of oral tongue squamous cell carcinoma identifies key variants and pathways linked to risk habits, HPV, clinical parameters and tumor recurrence. *F1000Res* **4:** 1215

Krishnan NM, Dhas K, Nair J, Palve V, Bagwan J, Siddappa G, Suresh A, Kekatpure VD, Kuriakose MA, Panda B (2016) A Minimal DNA Methylation Signature in Oral Tongue Squamous Cell Carcinoma Links Altered Methylation with Tumor Attributes. *Mol Cancer Res* **14**(9)**:** 805-19

Kuchnio A, Dewerchin M, Carmeliet P (2015) The PHD2 oxygen sensor paves the way to

metastasis. *Oncotarget* **6**(34)**:** 35149-50

Kurashina K, Yamashita Y, Ueno T, Koinuma K, Ohashi J, Horie H, Miyakura Y, Hamada T, Haruta H, Hatanaka H, Soda M, Choi YL, Takada S, Yasuda Y, Nagai H, Mano H (2008) Chromosome copy number analysis in screening for prognosis-related genomic regions in colorectal carcinoma. *Cancer Sci* **99**(9)**:** 1835-40

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* **12:** 323

Martinez-Outschoorn UE, Lin Z, Whitaker-Menezes D, Howell A, Sotgia F, Lisanti MP (2012) Ketone body utilization drives tumor growth and metastasis. *Cell Cycle* **11**(21)**:** 3964-71

Mishra A, Meherotra R (2014) Head and neck cancer: global burden and regional trends in India. *Asian Pacific journal of cancer prevention : APJCP* **15**(2)**:** 537-50

Mosca E, Alfieri R, Merelli I, Viti F, Calabria A, Milanesi L (2010) A multilevel data integration resource for breast cancer study. *BMC systems biology* **4:** 76

Mostovich LA, Prudnikova TY, Kondratov AG, Loginova D, Vavilov PV, Rykova VI, Sidorov SV, Pavlova TV, Kashuba VI, Zabarovsky ER, Grigorieva EV (2011) Integrin alpha9 (ITGA9) expression and epigenetic silencing in human breast tumors. *Cell Adh Migr* **5**(5)**:** 395-401

Murase H, Inokuchi M, Takagi Y, Kato K, Kojima K, Sugihara K (2014) Prognostic significance of the co-overexpression of fibroblast growth factor receptors 1, 2 and 4 in gastric cancer. *Mol Clin Oncol* **2**(4)**:** 509-517

Nakamura Y, Furukawa Y (2003) Method for treating or preventing metastasis of colorectal cancers: Google Patents. US20060111314

Nawaz I, Hu LF, Du ZM, Moumad K, Ignatyev I, Pavlova TV, Kashuba V, Almgren M, Zabarovsky ER, Ernberg I (2015) Integrin alpha9 gene promoter is hypermethylated and downregulated in nasopharyngeal carcinoma. *Oncotarget* **6**(31)**:** 31493-507

Netanely D, Avraham A, Ben-Baruch A, Evron E, Shamir R (2016) Expression and methylation patterns partition luminal-A breast tumors into distinct prognostic subgroups. *Breast cancer research : BCR* **18**(1)**:** 74

Pastuszak-Lewandoska D, Kordiak J, Antczak A, Migdalska-Sek M, Czarnecka KH, Gorski P, Nawrot E, Kiszalkiewicz JM, Domanska-Senderowska D, Brzezianska-Lasota E (2016) Expression level and methylation status of three tumor suppressor genes, DLEC1, ITGA9 and MLH1, in non-small cell lung cancer. *Med Oncol* **33**(7)**:** 75

Pickering CR, Zhang J, Yoo SY, Bengtsson L, Moorthy S, Neskey DM, Zhao M, Ortega Alves MV, Chang K, Drummond J, Cortez E, Xie TX, Zhang D, Chung W, Issa JP, Zweidler-McKay PA, Wu X, El-Naggar AK, Weinstein JN, Wang J, Muzny DM, Gibbs RA, Wheeler DA, Myers JN, Frederick MJ (2013) Integrative genomic characterization of oral squamous cell carcinoma identifies frequent somatic drivers. *Cancer discovery* **3**(7)**:** 770-81

Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC (2013) A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC genomics* **14:** 293

Richards JA (2013) *Remote Sensing Digital Image Analysis: An Introduction*: Springer Science & Business Media

Ridge JA, Mehra R, Lango MN, Galloway T (2016) Head and neck tumors. In *Cancer management: a multidisciplinary approach.*, Haller DG, Wagman LD, Camphausen KA, Hoskins WJ (eds), Chapter 2.

Scaife CL, Shea J, Emerson L, Boucher K, Firpo MA, Beckerle MC, Mulvihill SJ (2010) Prognostic significance of PINCH signalling in human pancreatic ductal adenocarcinoma. *HPB (Oxford)* **12**(5)**:** 352-8

Shi DB, Wang YW, Xing AY, Gao JW, Zhang H, Guo XY, Gao P (2015) C/EBPalpha-induced miR-100 expression suppresses tumor metastasis and growth by targeting ZBTB7A in gastric cancer. *Cancer Lett* **369**(2)**:** 376-85

Shim HJ, Shin MH, Kim HN, Kim JH, Hwang JE, Bae WK, Chung IJ, Cho SH (2016) The Prognostic Significance of FGFR4 Gly388 Polymorphism in Esophageal Squamous Cell Carcinoma after Concurrent Chemoradiotherapy. *Cancer Res Treat* **48**(1)**:** 71-9

Tandefelt DG, Boormans JL, van der Korput HA, Jenster GW, Trapman J (2013) A 36-gene signature predicts clinical progression in a subgroup of ERG-positive prostate cancers. *European urology* **64**(6)**:** 941-950

Thingholm LB, Andersen L, Makalic E, Southey MC, Thomassen M, Hansen LL (2016) Strategies for Integrated Analysis of Genetic, Epigenetic, and Gene Expression Variation in Cancer: Addressing the Challenges. *Frontiers in genetics* **7:** 2

Tsai K-CK, Chi-Rong L, SU J-MJ (2013) Molecular markers for prognostically predicting prostate cancer, method and kit thereof: Google Patents. US20130331281

Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Ponten F (2015) Proteomics. Tissue-based map of the human proteome. *Science* **347**(6220)**:** 1260419

Vauthey JN, Zimmitti G, Kopetz SE, Shindoh J, Chen SS, Andreou A, Curley SA, Aloia TA, Maru DM (2013) RAS mutation status predicts survival and patterns of recurrence in patients undergoing hepatectomy for colorectal liver metastases. *Ann Surg* **258**(4)**:** 619-26; discussion 626-7

Wang L, Song G, Tan W, Qi M, Zhang L, Chan J, Yu J, Han J, Han B (2015) MiR-573 inhibits prostate cancer metastasis by regulating epithelial-mesenchymal transition. *Oncotarget* **6**(34)**:** 35978-90

Wang Y, Zhang Y (2007) Methods of predicting distant metastasis of lymph node-negative primary breast cancer using biological pathway gene expression analysis: Google Patents. US20080182246

Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nature genetics* **45**(10)**:** 1113-20

Zhang F, Zhang X, Li M, Chen P, Zhang B, Guo H, Cao W, Wei X, Cao X, Hao X, Zhang N (2010) mTOR complex component Rictor interacts with PKCzeta and regulates cancer cell metastasis. *Cancer Res* **70**(22)**:** 9360-70

Zhang J, Na S, Liu C, Pan S, Cai J, Qiu J (2016) MicroRNA-125b suppresses the epithelial-mesenchymal transition and cell invasion by targeting ITGA9 in melanoma. *Tumour Biol* **37**(5)**:** 5941-9

Zhang YZ, Zhang LH, Gao Y, Li CH, Jia SQ, Liu N, Cheng F, Niu DY, Cho WC, Ji JF, Zeng CQ (2011) Discovery and validation of prognostic markers in gastric cancer by genome-wide expression profiling. *World J Gastroenterol* **17**(13)**:** 1710-7

Zhao Z, Han FH, Yang SB, Hua LX, Wu JH, Zhan WH (2015) Loss of stromal caveolin-1 expression in colorectal cancer predicts poor survival. *World J Gastroenterol* **21**(4)**:** 1140-7

Table 1: Sample cohorts used in the discovery and validation modules of *CAFE MOCHA* for discovering metastatic/recurrent signature MR44 in head and neck squamous cell carcinoma (HNSCC).

Discovery			
Stage 1	Source	TCGA HNSCC (http://tinyurl.com/jhmg9b9), data available on all the 4 event types (mut, CNV, exp, meth) on tumor:matched normal samples.	
	Total Number of tumors	434	
	Metastatic/Recurrent	83	
	Tumors without metastasis/recurrence	351	
Stage 2	Source	Data on at least one event type available from the same matched tumor:normal sample.	
		TCGA HNSCC	OTSCC (Krishnan *et al*, 2015; Krishnan *et al*, 2016)
	Total number of tumors	42	37
	Metastatic/Recurrent	4	15
	Tumors without metastasis/recurrence	38	22
	Adjacent normal tissue	0	37
Validation			
	Source	OTSCC (Krishnan *et al*, 2015; Krishnan *et al*, 2016) (matched tumor:normal data on all event types from the same tumors)	
	Total number of tumors	18	
	Metastatic/Recurrent	3	
	Tumors without metastasis/recurrence	15	
	Adjacent normal tissue	18	

741

742

743 **Figure Legends**

744 Figure 1: *CAFE MOCHA* application workflow and graphical-user-interface. A. Discovery,

745 and Prediction modules. B. Integrated analyses.

746

747 Figure 2: Complete specificity margin based clustering (CSMBC) algorithm for quantitative

748 and coupled events.

749 Clustering of quantitative and coupled events, for three clinical sub-categories *i*, *j* and *k* is

750 demonstrated here. Panel A: linking of quantitative events such as expression or

751 methylation to clinical sub-categories *i*, *j* and *k* (shades of blue). Boundaries are

752 determined in a supervised manner, as the minimum and maximum limits of the

753 quantitative ranges for each clinical sub-category. Overlaps are estimated as per the

754 equations illustrated in the figure. Samples whose expression or methylation values lie

755 outside the regions of overlap, 100% specific to a clinical sub-category are factored in

756 towards the sensitivity of that gene for that clinical sub-category. Panel B: clinical linking of

757 coupled events such as fmeth (shades of orange) resulting from a combination of two

758 quantitative events (expr – shades of blue and meth – shades of purple). The expression

759 event is mapped along the first dimension and the methylation event is mapped along the

760 second, and the samples, which observe both, expression and methylation, boundaries,

761 contribute to the sensitivity of that fmeth event for that clinical sub-category. Panel C:

762 clinical linking of coupled events such as fcnv (shades of dark orange) resulting from a

763 combination of one quantitative event (expression – shades of blue) and another

764 categorical event (CNV – shades of green). Here, the fcnv event linked to a clinical sub-

765 category is determined by both, expression boundaries for a clinical sub-group, and

766 presence of CNVs for the same samples in that clinical sub-group.

Figure 3: Discovery heatmap (A) and per-event validation sensitivity (B) of 44-gene integrated signature (MR44) for distant metastasis and recurrence in the HNSCC discovery set.

An integrated signature associated with metastasis/recurrence was derived from combining six event types (mut: red, fmeth: dark orange, fcnv: orange, cnv: green, expr: blue and meth: purple). Cumulative sample frequency (%) for individual event types is represented as a histogram. The per-event validation sensitivity is represented as bubbles, where the bubble size is proportional to the sample frequency of that event in the validation cohort.

Figure 4: Comparison of detection sensitivities and total numbers of events required achieving the sensitivity in one (A), two (B), three (C), four (D), five (E) and all six (F) event types. Utilizing all the six events (F) shows the power of integration both on sensitivity of detection (black dot) and the number of events (colored bars for all the six individual event types) required to attain the sensitivity.

Supplementary Data Legend

Supplementary Table S1: TCGA HNSCC tumor samples (*n* = 434) and clinical attributes used for the discovery of MR44.

Supplementary Table S2: Oral tongue squamous cell carcinoma (OTSCC) and TCGA HNSCC samples lacking cross-platform overlap, and their clinical attributes used for confirmation of MR44 discovery.

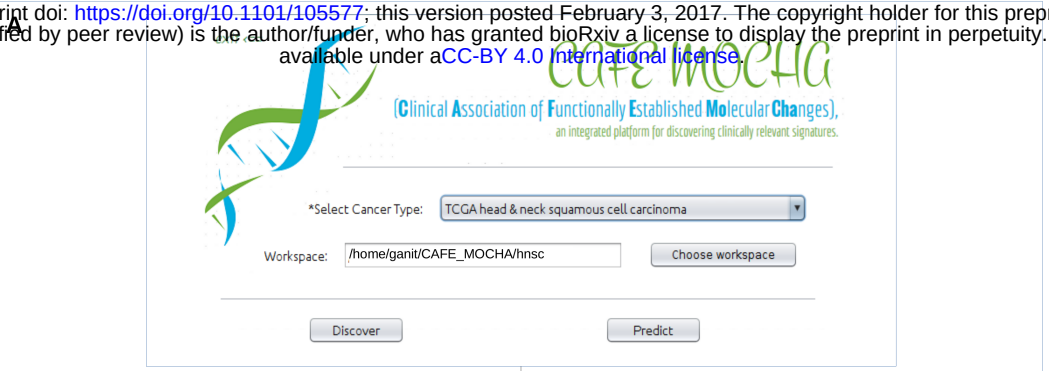
Supplementary Table S3: Oral tongue squamous cell carcinoma (OTSCC) (*n* = 18) with all

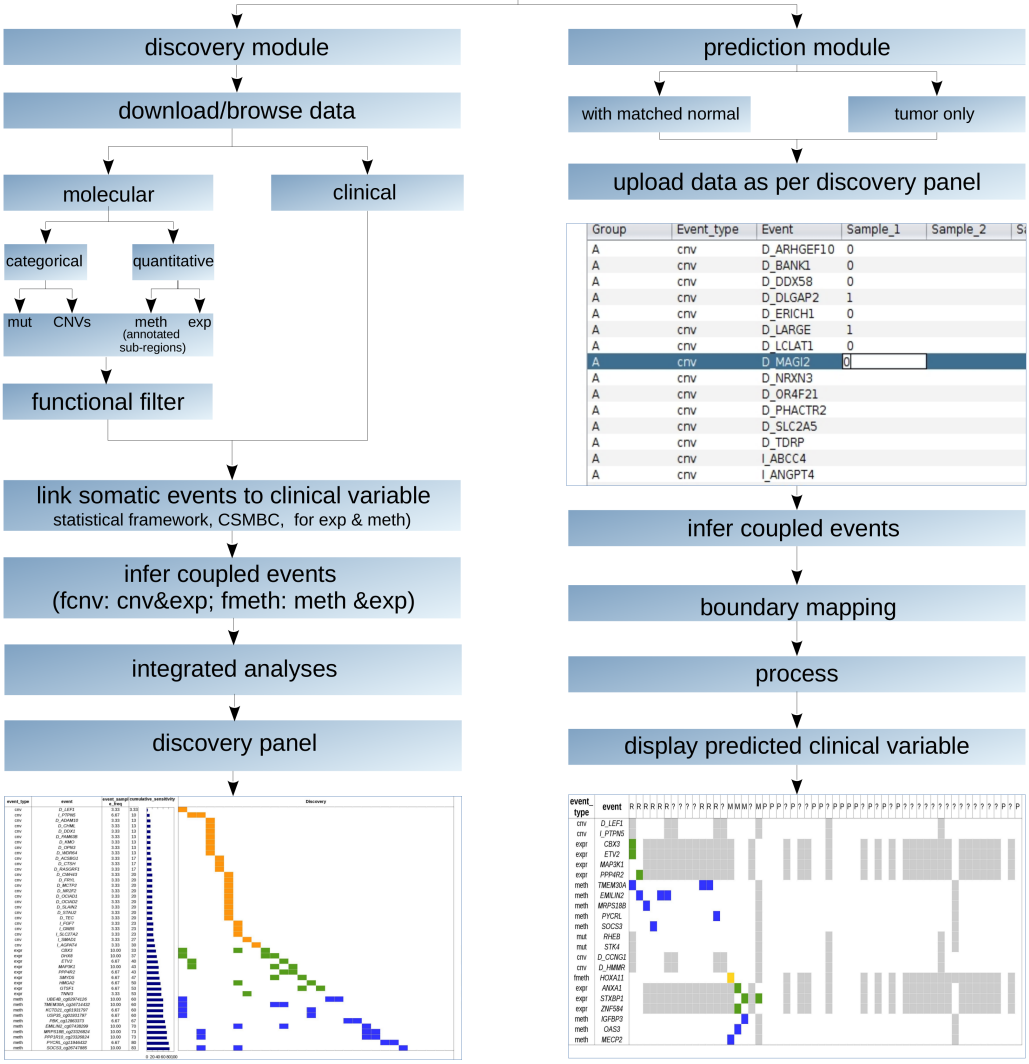
four events assayed within the same tumor, and their clinical attributes used for the validation.

Supplementary Table S4: Discovered events associated with Metastases and Recurrence in HNSCC, pathways mapped, confirmation and validation status.

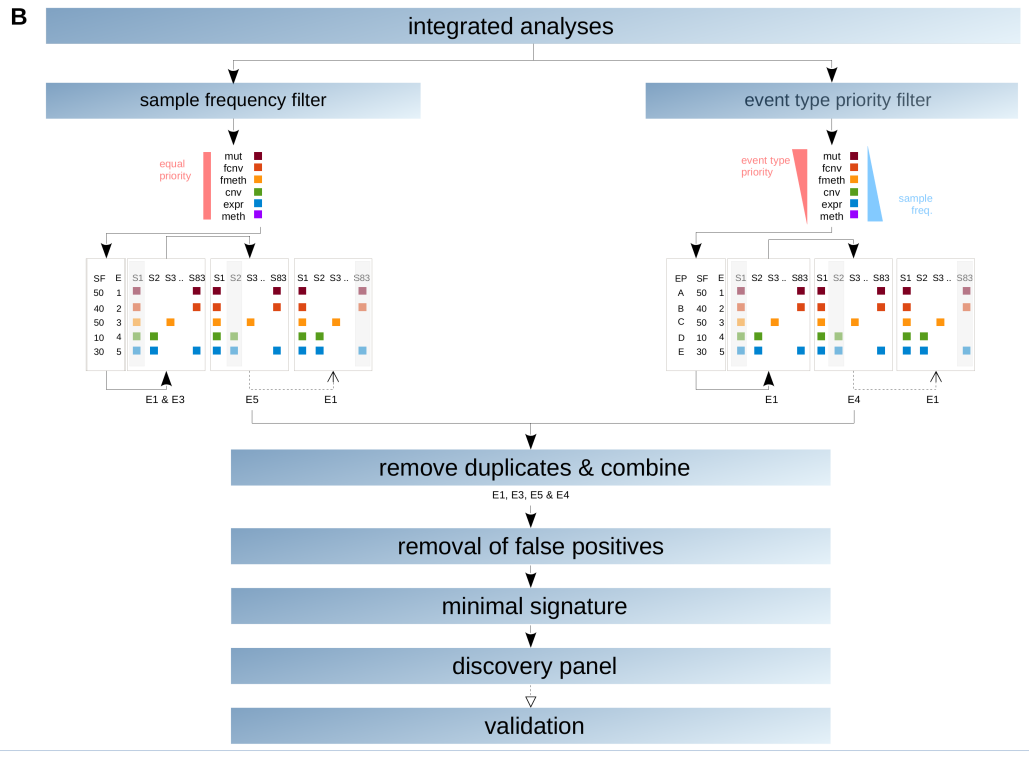
Supplementary Figure S1: Events associated with distant metastasis and recurrence in HNSCC with single-event or multi-event signatures, selected in individual event type (mut: red, fmeth: d orange, fcnv: orange, cnv: green, expr: blue and meth: purple) and various integrated analyses, combining different numbers of event types.

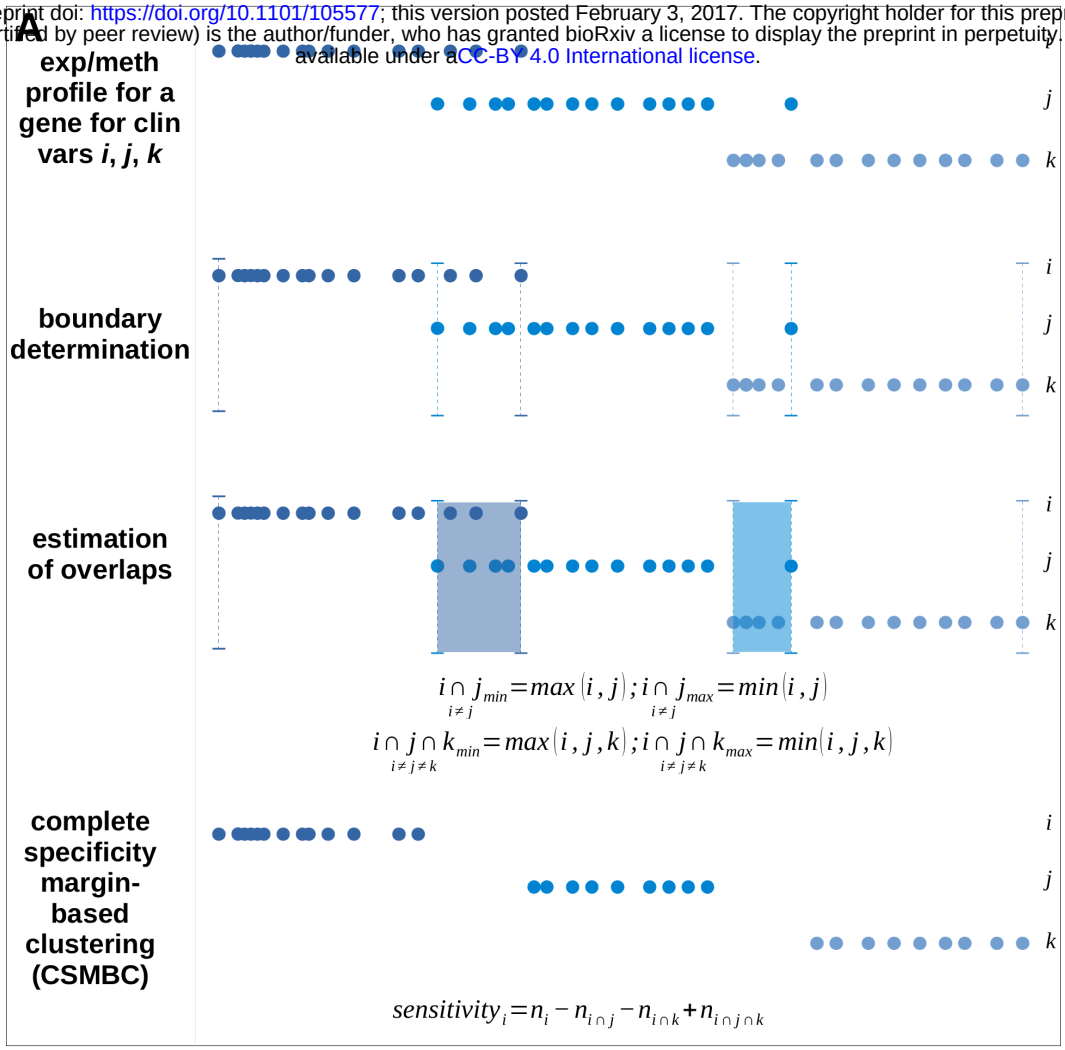
A

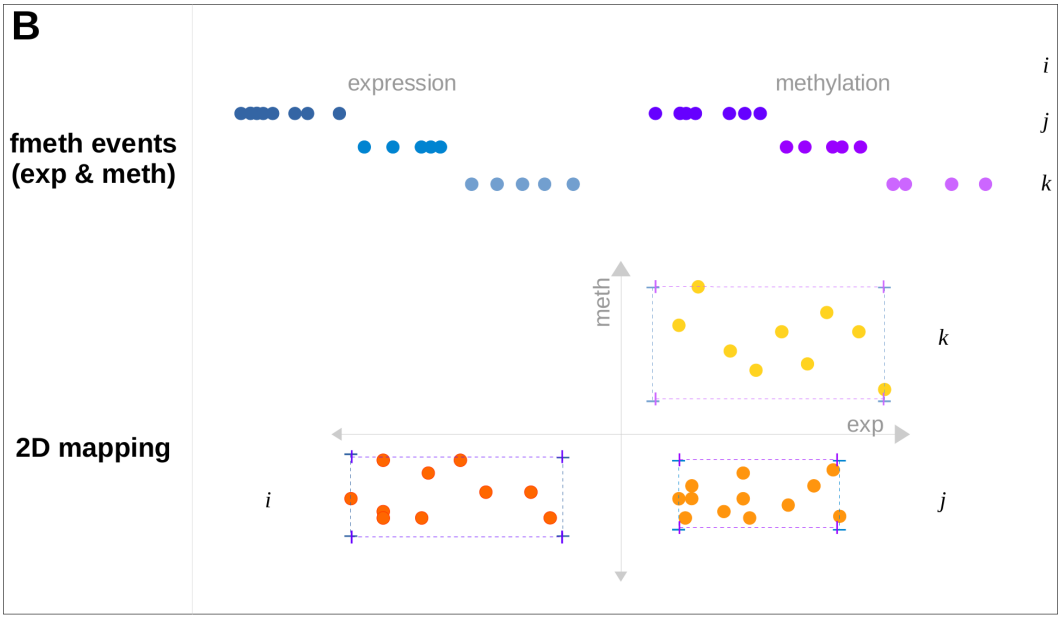


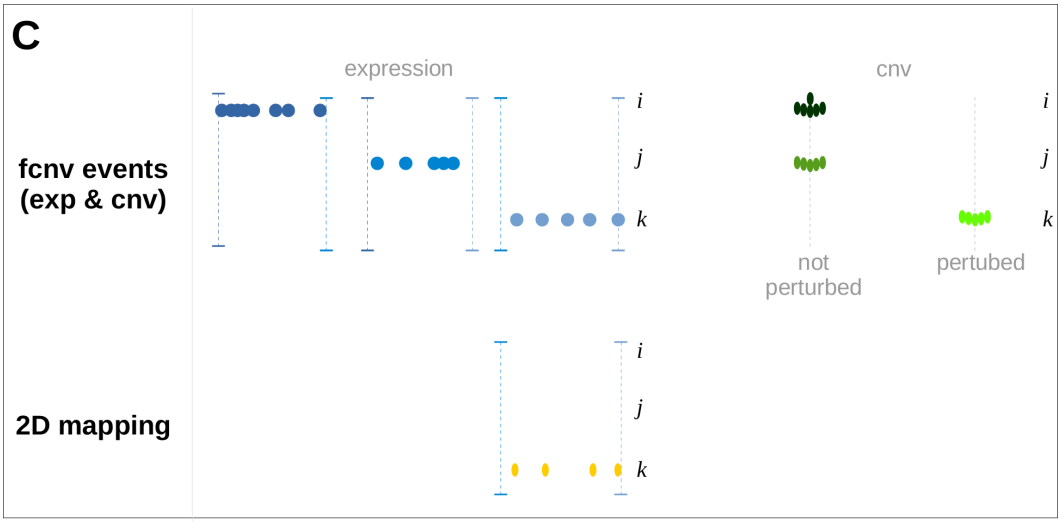


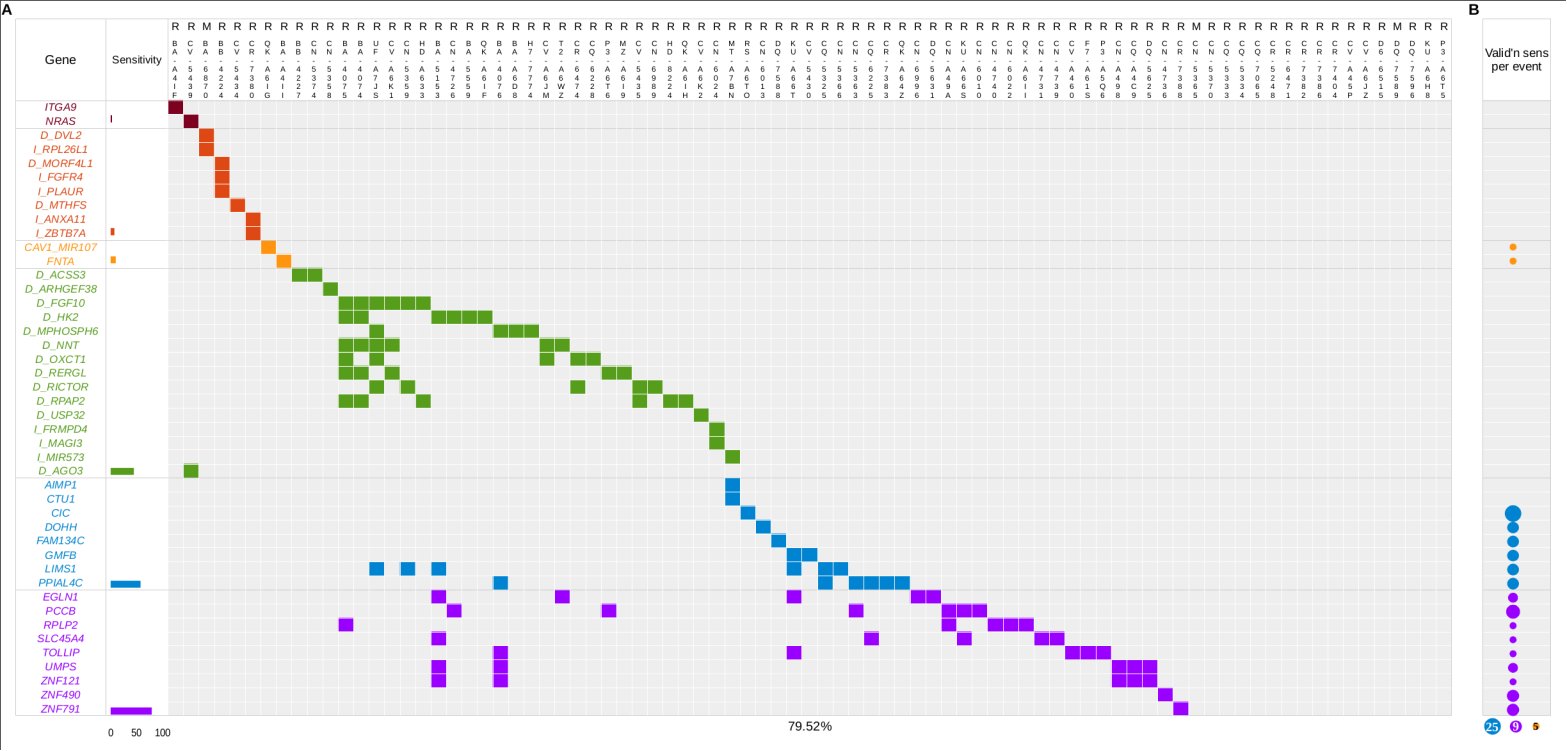
B

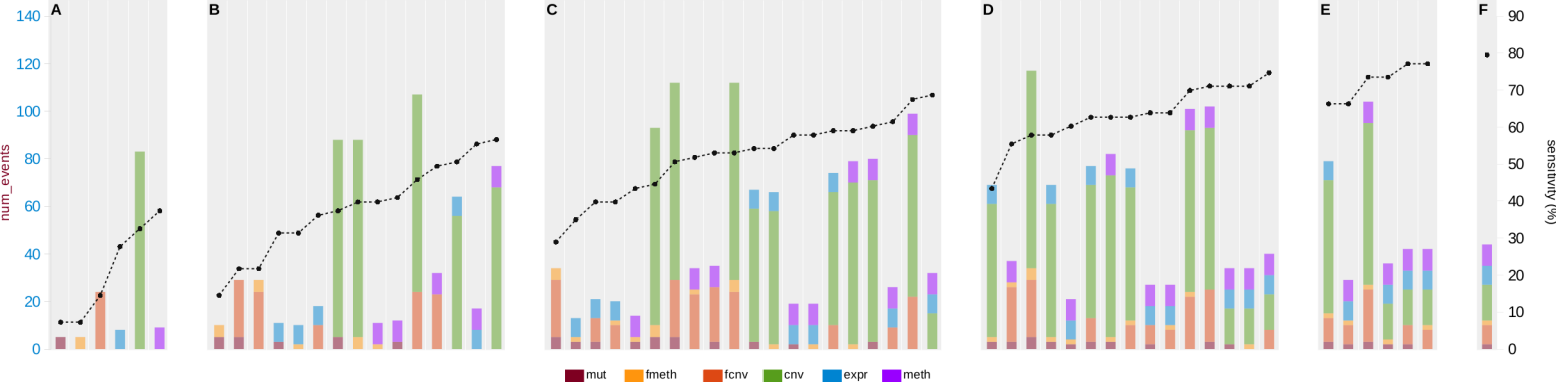




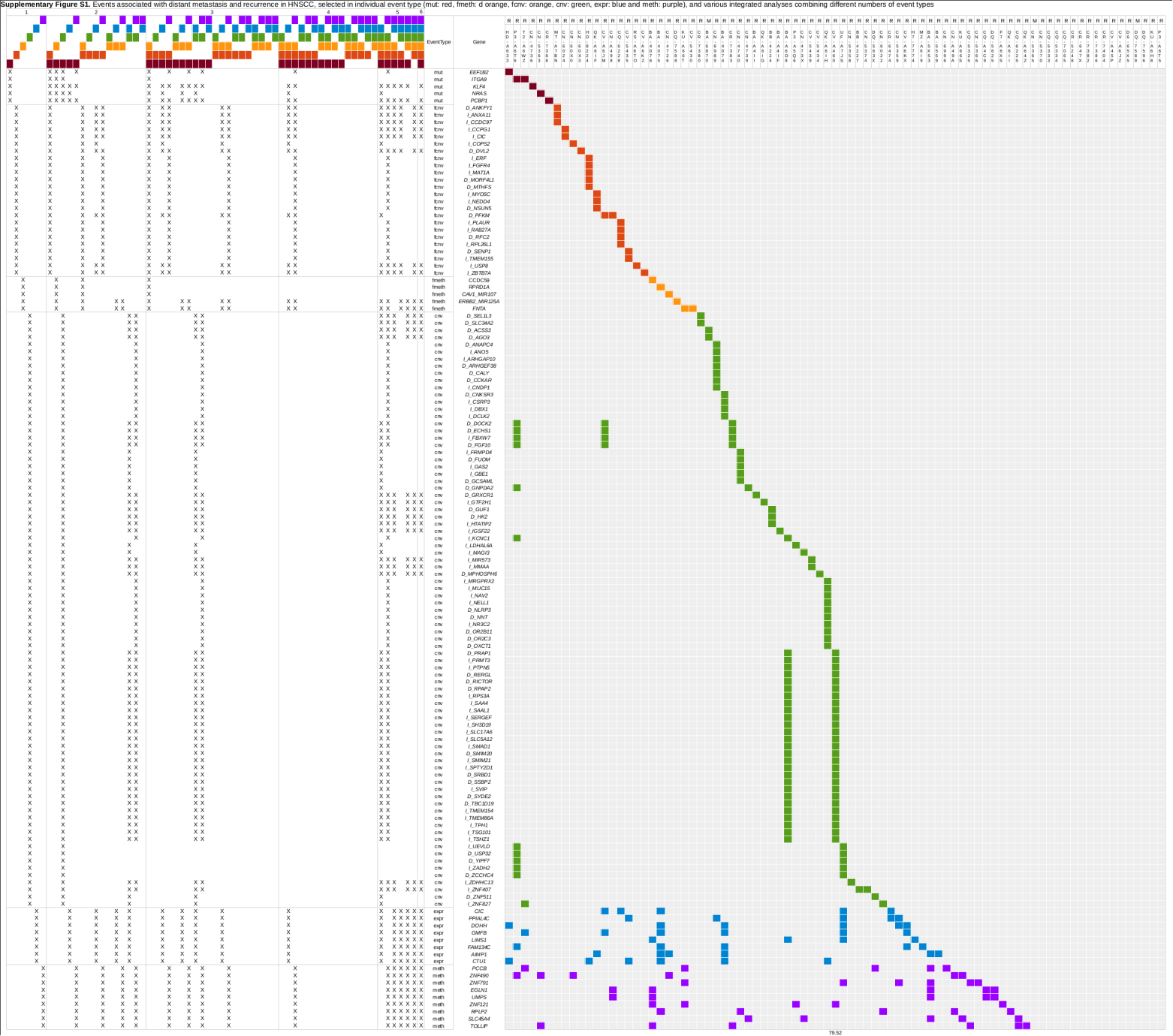





A
B
Gene
Sensitivity
ITGA9
NRAS
D_DVL2
I_RPL26L1
D_MORF4L1
I_FGFR4
I_PLAUR
D_MTHFS
I_ANXA11
I_ZBTB7A
CAV1_MIR107
FNTA
D_ACSS3
D_ARHGEF38
D_FGF10
D_HK2
D_MPHOSPH6
D_NNT
D_OXCT1
D_RERGL
D_RICTOR
D_RPAP2
D_USP32
I_FRMPD4
I_MAGI3
I_MIR573
D_AGO3
AIMP1
CTU1
CIC
DOHH
FAM134C
GMFB
LIMS1
PPIAL4C
EGLN1
PCCB
RPLP2
SLC45A4
TOLLIP
UMPS
ZNF121
ZNF490
ZNF791
0
50
100
79.52%
Valid'n sens per event
25
9
5


A
B
C
D
E
F
num_events
sensitivity (%)
0
20
40
60
80
100
120
140
0
10
20
30
40
50
60
70
80
90
mut
fmeth
fcnv
cnv
expr
meth

Supplementary Figure S1. Events associated with distant metastasis and recurrence in HNSCC, selected in individual event type (mut: red, fmeth: d orange, fcnv: orange, cnv: green, expr: blue and meth: purple), and various integrated analyses combining different numbers of event types



Supplementary Table S1: TCGA HNSCC samples and clinical attributes used for discovery.

SampleID	R (Loco-regional recurrence); M1(metastasis); M0 (Non-metastatic/Non-recurrent) tumors; N (Solid Tissue Normals).
TCGA-BA-4074-01	R
TCGA-BA-4075-01	R
TCGA-BA-4076-01	R
TCGA-BA-4077-01	M0
TCGA-BA-4078-01	M0
TCGA-BA-5149-01	M0
TCGA-BA-5151-01	M0
TCGA-BA-5152-01	M0
TCGA-BA-5153-01	R
TCGA-BA-5555-01	M0
TCGA-BA-5556-01	M0
TCGA-BA-5557-01	M0
TCGA-BA-5558-01	M0
TCGA-BA-5559-01	R
TCGA-BA-6868-01	M0
TCGA-BA-6869-01	M0
TCGA-BA-6870-01	M1
TCGA-BA-6871-01	M0
TCGA-BA-6872-01	M0
TCGA-BA-6873-01	M0
TCGA-BA-7269-01	M0
TCGA-BA-A4IF-01	R
TCGA-BA-A4IH-01	M0

TCGA-BA-A4II-01	R
TCGA-BA-A6D8-01	R
TCGA-BA-A6DA-01	M0
TCGA-BA-A6DB-01	M0
TCGA-BA-A6DD-01	M0
TCGA-BA-A6DE-01	M0
TCGA-BA-A6DI-01	M0
TCGA-BA-A6DJ-01	M0
TCGA-BA-A6DL-01	M0
TCGA-BB-4217-01	M0
TCGA-BB-4223-01	M0
TCGA-BB-4224-01	R
TCGA-BB-4225-01	M0
TCGA-BB-4227-01	R
TCGA-BB-4228-01	M0
TCGA-BB-7861-01	M0
TCGA-BB-7862-01	M0
TCGA-BB-7863-01	M0
TCGA-BB-7864-01	M0
TCGA-BB-7870-01	M0
TCGA-BB-7871-01	M0
TCGA-BB-7872-01	M0
TCGA-C9-A47Z-01	M0
TCGA-C9-A480-01	M0
TCGA-CN-4723-01	M0
TCGA-CN-4725-01	M0
TCGA-CN-4726-01	R

TCGA-CN-4727-01	M0
TCGA-CN-4728-01	M0
TCGA-CN-4729-01	M0
TCGA-CN-4730-01	M0
TCGA-CN-4731-01	R
TCGA-CN-4733-01	M0
TCGA-CN-4735-01	M0
TCGA-CN-4736-01	R
TCGA-CN-4737-01	M0
TCGA-CN-4738-01	M0
TCGA-CN-4739-01	R
TCGA-CN-4740-01	R
TCGA-CN-4741-01	M0
TCGA-CN-4742-01	M0
TCGA-CN-5355-01	M0
TCGA-CN-5356-01	M0
TCGA-CN-5358-01	R
TCGA-CN-5359-01	R
TCGA-CN-5360-01	M0
TCGA-CN-5363-01	R
TCGA-CN-5364-01	M0
TCGA-CN-5365-01	M1
TCGA-CN-5366-01	R
TCGA-CN-5367-01	M0
TCGA-CN-5369-01	M0
TCGA-CN-5370-01	R
TCGA-CN-5373-01	M0

TCGA-CN-5374-01	R
TCGA-CN-6010-01	R
TCGA-CN-6011-01	M0
TCGA-CN-6012-01	M0
TCGA-CN-6013-01	R
TCGA-CN-6016-01	M0
TCGA-CN-6017-01	M0
TCGA-CN-6018-01	M0
TCGA-CN-6019-01	M0
TCGA-CN-6020-01	M0
TCGA-CN-6021-01	M0
TCGA-CN-6022-01	R
TCGA-CN-6023-01	M0
TCGA-CN-6024-01	R
TCGA-CN-6988-01	M0
TCGA-CN-6989-01	R
TCGA-CN-6992-01	M0
TCGA-CN-6994-01	M0
TCGA-CN-6995-01	M0
TCGA-CN-6996-01	R
TCGA-CN-6997-01	M0
TCGA-CN-6998-01	M0
TCGA-CN-A498-01	R
TCGA-CN-A49A-01	R
TCGA-CN-A642-01	M0
TCGA-CN-A6V3-01	M0
TCGA-CQ-5323-01	M0

TCGA-CQ-5324-01	M0
TCGA-CQ-5325-01	R
TCGA-CQ-5326-01	M0
TCGA-CQ-5327-01	M0
TCGA-CQ-5329-01	M0
TCGA-CQ-5330-01	M0
TCGA-CQ-5331-01	M0
TCGA-CQ-5332-01	M0
TCGA-CQ-5333-01	R
TCGA-CQ-5334-01	R
TCGA-CQ-6218-01	M0
TCGA-CQ-6220-01	M0
TCGA-CQ-6221-01	M0
TCGA-CQ-6223-01	M0
TCGA-CQ-6224-01	M0
TCGA-CQ-6225-01	R
TCGA-CQ-6227-01	M0
TCGA-CQ-6228-01	R
TCGA-CQ-6229-01	M0
TCGA-CQ-7063-01	M0
TCGA-CQ-7065-01	R
TCGA-CQ-7067-01	M0
TCGA-CQ-7068-01	M0
TCGA-CQ-7069-01	M0
TCGA-CQ-7071-01	M0
TCGA-CQ-7072-01	M0
TCGA-CQ-A4C6-01	M0

TCGA-CQ-A4C7-01	M0
TCGA-CQ-A4C9-01	R
TCGA-CQ-A4CB-01	M0
TCGA-CQ-A4CD-01	M0
TCGA-CQ-A4CE-01	M0
TCGA-CQ-A4CG-01	M0
TCGA-CQ-A4CH-01	M0
TCGA-CR-5243-01	M0
TCGA-CR-5247-01	M0
TCGA-CR-5248-01	R
TCGA-CR-5249-01	M0
TCGA-CR-6467-01	M0
TCGA-CR-6470-01	M0
TCGA-CR-6471-01	R
TCGA-CR-6472-01	M0
TCGA-CR-6473-01	M0
TCGA-CR-6474-01	R
TCGA-CR-6477-01	M0
TCGA-CR-6478-01	M0
TCGA-CR-6480-01	M0
TCGA-CR-6481-01	M0
TCGA-CR-6482-01	M0
TCGA-CR-6484-01	M0
TCGA-CR-6487-01	M0
TCGA-CR-6488-01	M0
TCGA-CR-6491-01	M0
TCGA-CR-6492-01	M0

TCGA-CR-6493-01	M0
TCGA-CR-7364-01	M0
TCGA-CR-7365-01	M0
TCGA-CR-7367-01	M0
TCGA-CR-7368-01	M0
TCGA-CR-7369-01	M0
TCGA-CR-7370-01	M0
TCGA-CR-7371-01	M0
TCGA-CR-7372-01	M0
TCGA-CR-7373-01	M0
TCGA-CR-7374-01	M0
TCGA-CR-7376-01	M0
TCGA-CR-7377-01	M0
TCGA-CR-7379-01	M0
TCGA-CR-7380-01	R
TCGA-CR-7382-01	R
TCGA-CR-7383-01	R
TCGA-CR-7385-01	M0
TCGA-CR-7386-01	R
TCGA-CR-7388-01	R
TCGA-CR-7389-01	M0
TCGA-CR-7390-01	M0
TCGA-CR-7391-01	M0
TCGA-CR-7392-01	M0
TCGA-CR-7393-01	M0
TCGA-CR-7394-01	M0
TCGA-CR-7395-01	M0

TCGA-CR-7397-01	M0
TCGA-CR-7398-01	M0
TCGA-CR-7399-01	M0
TCGA-CR-7401-01	M0
TCGA-CR-7402-01	M0
TCGA-CR-7404-01	R
TCGA-CV-5430-01	R
TCGA-CV-5432-01	M0
TCGA-CV-5434-01	R
TCGA-CV-5435-01	R
TCGA-CV-5436-01	M0
TCGA-CV-5439-01	R
TCGA-CV-5440-01	M0
TCGA-CV-5441-01	M0
TCGA-CV-5442-01	M0
TCGA-CV-5443-01	M0
TCGA-CV-5444-01	M0
TCGA-CV-5966-01	M0
TCGA-CV-5970-01	M0
TCGA-CV-5971-01	M0
TCGA-CV-5973-01	M0
TCGA-CV-5976-01	M0
TCGA-CV-5977-01	M0
TCGA-CV-5978-01	M0
TCGA-CV-5979-01	M0
TCGA-CV-6003-01	M0
TCGA-CV-6433-01	M0

TCGA-CV-6436-01	M0
TCGA-CV-6441-01	M0
TCGA-CV-6933-01	M0
TCGA-CV-6934-01	M0
TCGA-CV-6935-01	M0
TCGA-CV-6936-01	M0
TCGA-CV-6937-01	M0
TCGA-CV-6938-01	M0
TCGA-CV-6939-01	M0
TCGA-CV-6940-01	M0
TCGA-CV-6941-01	M0
TCGA-CV-6942-01	M0
TCGA-CV-6943-01	M0
TCGA-CV-6945-01	M0
TCGA-CV-6948-01	M0
TCGA-CV-6950-01	M0
TCGA-CV-6951-01	M0
TCGA-CV-6952-01	M0
TCGA-CV-6953-01	M0
TCGA-CV-6954-01	M0
TCGA-CV-6955-01	M0
TCGA-CV-6956-01	M0
TCGA-CV-6959-01	M0
TCGA-CV-6960-01	M0
TCGA-CV-6961-01	M0
TCGA-CV-6962-01	M0
TCGA-CV-7089-01	M0

TCGA-CV-7090-01	M0
TCGA-CV-7091-01	M0
TCGA-CV-7095-01	M0
TCGA-CV-7097-01	M0
TCGA-CV-7099-01	M0
TCGA-CV-7100-01	M0
TCGA-CV-7101-01	M0
TCGA-CV-7102-01	M0
TCGA-CV-7103-01	M0
TCGA-CV-7104-01	M0
TCGA-CV-7177-01	M0
TCGA-CV-7178-01	M0
TCGA-CV-7180-01	M0
TCGA-CV-7183-01	M0
TCGA-CV-7235-01	M0
TCGA-CV-7236-01	M0
TCGA-CV-7238-01	M0
TCGA-CV-7242-01	M0
TCGA-CV-7243-01	M0
TCGA-CV-7245-01	M0
TCGA-CV-7247-01	M0
TCGA-CV-7248-01	M0
TCGA-CV-7250-01	M0
TCGA-CV-7252-01	M0
TCGA-CV-7253-01	M0
TCGA-CV-7254-01	M0
TCGA-CV-7255-01	M0

TCGA-CV-7261-01	M0
TCGA-CV-7263-01	M0
TCGA-CV-7406-01	M0
TCGA-CV-7407-01	M0
TCGA-CV-7409-01	M0
TCGA-CV-7410-01	M0
TCGA-CV-7411-01	M0
TCGA-CV-7413-01	M0
TCGA-CV-7414-01	M0
TCGA-CV-7415-01	M0
TCGA-CV-7416-01	M0
TCGA-CV-7418-01	M0
TCGA-CV-7421-01	M0
TCGA-CV-7422-01	M0
TCGA-CV-7423-01	M0
TCGA-CV-7424-01	M0
TCGA-CV-7425-01	M0
TCGA-CV-7427-01	M0
TCGA-CV-7429-01	M0
TCGA-CV-7430-01	M0
TCGA-CV-7432-01	M0
TCGA-CV-7433-01	M0
TCGA-CV-7434-01	M0
TCGA-CV-7435-01	M0
TCGA-CV-7437-01	M0
TCGA-CV-7438-01	M0
TCGA-CV-7440-01	M0

TCGA-CV-A45O-01	M0
TCGA-CV-A45P-01	R
TCGA-CV-A45Q-01	M0
TCGA-CV-A45R-01	M0
TCGA-CV-A45T-01	M0
TCGA-CV-A45U-01	M0
TCGA-CV-A45V-01	M0
TCGA-CV-A45W-01	M0
TCGA-CV-A45X-01	M0
TCGA-CV-A45Y-01	M0
TCGA-CV-A45Z-01	M0
TCGA-CV-A460-01	R
TCGA-CV-A461-01	M0
TCGA-CV-A463-01	M0
TCGA-CV-A464-01	M0
TCGA-CV-A465-01	M0
TCGA-CV-A468-01	M0
TCGA-CV-A6JD-01	M0
TCGA-CV-A6JE-01	M0
TCGA-CV-A6JM-01	R
TCGA-CV-A6JN-01	M0
TCGA-CV-A6JO-01	M0
TCGA-CV-A6JT-01	M0
TCGA-CV-A6JU-01	M0
TCGA-CV-A6JY-01	M0
TCGA-CV-A6JZ-01	R
TCGA-CV-A6K0-01	M0

TCGA-CV-A6K1-01	R
TCGA-CV-A6K2-01	R
TCGA-CX-7082-01	M0
TCGA-CX-7085-01	M0
TCGA-CX-7086-01	M0
TCGA-CX-7219-01	M0
TCGA-D6-6515-01	R
TCGA-D6-6516-01	M0
TCGA-D6-6517-01	M0
TCGA-D6-6823-01	M0
TCGA-D6-6824-01	M0
TCGA-D6-6825-01	M0
TCGA-D6-6826-01	M0
TCGA-D6-6827-01	M0
TCGA-D6-8568-01	M0
TCGA-D6-8569-01	M0
TCGA-D6-A4Z9-01	M0
TCGA-D6-A4ZB-01	M0
TCGA-D6-A6EK-01	M0
TCGA-D6-A6EM-01	M0
TCGA-D6-A6EO-01	M0
TCGA-D6-A6EP-01	M0
TCGA-D6-A6EQ-01	M0
TCGA-D6-A6ES-01	M0
TCGA-D6-A74Q-01	M0
TCGA-DQ-5624-01	M0
TCGA-DQ-5625-01	R

TCGA-DQ-5629-01	M0
TCGA-DQ-5630-01	M0
TCGA-DQ-5631-01	R
TCGA-DQ-7588-01	R
TCGA-DQ-7589-01	M1
TCGA-DQ-7590-01	M0
TCGA-DQ-7591-01	M0
TCGA-DQ-7592-01	M0
TCGA-DQ-7593-01	M0
TCGA-DQ-7594-01	M0
TCGA-DQ-7595-01	M0
TCGA-DQ-7596-01	R
TCGA-F7-7848-01	M0
TCGA-F7-8489-01	M0
TCGA-F7-A50G-01	M0
TCGA-F7-A50I-01	M0
TCGA-F7-A50J-01	M0
TCGA-F7-A61S-01	R
TCGA-H7-7774-01	R
TCGA-H7-8501-01	M0
TCGA-H7-A6C4-01	M0
TCGA-HD-7229-01	M0
TCGA-HD-7753-01	M0
TCGA-HD-7754-01	M0
TCGA-HD-7831-01	M0
TCGA-HD-7832-01	M0
TCGA-HD-7917-01	M0

TCGA-HD-8224-01	R
TCGA-HD-8314-01	M0
TCGA-HD-A633-01	R
TCGA-HL-7533-01	M0
TCGA-IQ-7630-01	M0
TCGA-IQ-7631-01	M0
TCGA-IQ-7632-01	M0
TCGA-IQ-A61I-01	M0
TCGA-IQ-A61J-01	M0
TCGA-KU-A66S-01	R
TCGA-KU-A66T-01	R
TCGA-KU-A6H7-01	M0
TCGA-KU-A6H8-01	R
TCGA-MT-A51W-01	M0
TCGA-MT-A51X-01	M0
TCGA-MT-A67D-01	M0
TCGA-MT-A7BN-01	R
TCGA-MZ-A5BI-01	M0
TCGA-MZ-A6I9-01	R
TCGA-P3-A5Q6-01	R
TCGA-P3-A6T5-01	R
TCGA-P3-A6T6-01	R
TCGA-QK-A64Z-01	R
TCGA-QK-A6IF-01	R
TCGA-QK-A6IG-01	R
TCGA-QK-A6IH-01	R
TCGA-QK-A6II-01	R

TCGA-QK-A6IJ-01	M0
TCGA-QK-A6V9-01	M0
TCGA-QK-A6VB-01	M0
TCGA-QK-A6VC-01	M0
TCGA-RS-A6TO-01	R
TCGA-RS-A6TP-01	M0
TCGA-T2-A6WX-01	M0
TCGA-T2-A6WZ-01	R
TCGA-T2-A6X0-01	M0
TCGA-T2-A6X2-01	M0
TCGA-TN-A7HI-01	M0
TCGA-TN-A7HJ-01	M0
TCGA-TN-A7HL-01	M0
TCGA-UF-A718-01	M0
TCGA-UF-A719-01	M0
TCGA-UF-A71A-01	M0
TCGA-UF-A71B-01	M0
TCGA-UF-A71D-01	M0
TCGA-UF-A71E-01	M0
TCGA-UF-A7J9-01	M0
TCGA-UF-A7JA-01	M0
TCGA-UF-A7JC-01	M0
TCGA-UF-A7JD-01	M0
TCGA-UF-A7JF-01	M0
TCGA-UF-A7JH-01	M0
TCGA-UF-A7JJ-01	M0
TCGA-UF-A7JK-01	M0

TCGA-UF-A7JO-01	M0
TCGA-UF-A7JS-01	R
TCGA-UF-A7JT-01	M0
TCGA-UF-A7JV-01	M0
TCGA-WA-A7GZ-01	M0
TCGA-WA-A7H4-01	M0
TCGA-CV-5430-11	N
TCGA-CV-5431-11	N
TCGA-CV-5432-11	N
TCGA-CV-5434-11	N
TCGA-CV-5435-11	N
TCGA-CV-5436-11	N
TCGA-CV-5439-11	N
TCGA-CV-5440-11	N
TCGA-CV-5441-11	N
TCGA-CV-5442-11	N
TCGA-CV-5443-11	N
TCGA-CV-5444-11	N
TCGA-CV-5966-11	N
TCGA-CV-5970-11	N
TCGA-CV-5971-11	N
TCGA-CV-5973-11	N
TCGA-CV-5976-11	N
TCGA-CV-5977-11	N
TCGA-CV-5978-11	N
TCGA-CV-5979-11	N
TCGA-CV-6003-11	N

TCGA-CV-6433-11	N
TCGA-CV-6436-11	N
TCGA-CV-6441-11	N
TCGA-CV-6933-11	N
TCGA-CV-6934-11	N
TCGA-CV-6935-11	N
TCGA-CV-6936-11	N
TCGA-CV-6938-11	N
TCGA-CV-6939-11	N
TCGA-CV-6943-11	N
TCGA-CV-6951-11	N
TCGA-CV-6952-11	N
TCGA-CV-6953-11	N
TCGA-CV-6954-11	N
TCGA-CV-6955-11	N
TCGA-CV-6956-11	N
TCGA-CV-6959-11	N
TCGA-CV-6960-11	N
TCGA-CV-6961-11	N
TCGA-CV-6962-11	N
TCGA-CV-7089-11	N
TCGA-CV-7091-11	N
TCGA-CV-7097-11	N
TCGA-CV-7101-11	N
TCGA-CV-7103-11	N
TCGA-CV-7177-11	N
TCGA-CV-7178-11	N

TCGA-CV-7183-11	N
TCGA-CV-7235-11	N
TCGA-CV-7238-11	N
TCGA-CV-7242-11	N
TCGA-CV-7245-11	N
TCGA-CV-7250-11	N
TCGA-CV-7252-11	N
TCGA-CV-7255-11	N
TCGA-CV-7261-11	N
TCGA-CV-7263-11	N
TCGA-CV-7406-11	N
TCGA-CV-7416-11	N
TCGA-CV-7423-11	N
TCGA-CV-7424-11	N
TCGA-CV-7425-11	N
TCGA-CV-7432-11	N
TCGA-CV-7434-11	N
TCGA-CV-7437-11	N
TCGA-CV-7438-11	N
TCGA-CV-7440-11	N
TCGA-H7-A6C5-11	N
TCGA-HD-8635-11	N
TCGA-HD-A6HZ-11	N
TCGA-HD-A6I0-11	N
TCGA-WA-A7GZ-11	N

Supplementary Table S2: Oral tongue squamous cell carcinoma (OTSCC) and TCGA HNSC samples lacking cross-platform overlap, and their clinical attributes used for validation.

SampleID	R (Loco-regional recurrence); M1(metastasis); M0 (Non-metastatic/Non-recurrent) tumors; N (Solid Tissue Normals)
OT10_T	M0
OT13_T	M0
OT15_T	R
OT16_T	R
OT18_T	M1
OT20_T	R
OT21_T	R
OT23_T	M0
OT25_T	M0
OT27_T	M0
OT29_T	M0
OT30_T	M0
OT31_T	M0
OT32_T	M0
OT33_T	M0
OT34_T	R
OT35_T	M0
OT36_T	R
OT37_T	M0
OT38_T	R
OT39_T	M0
OT4_T	R

OT40_T	R
OT41_T	M0
OT42_T	M0
OT43_T	M0
OT45_T	M0
OT46_T	M0
OT47_T	M0
OT48_T	R
OT49_T	R
OT5_T	R
OT50_T	M1
OT53_T	M0
OT55_T	R
OT7_T	M0
OT9_T	M0
TCGA-BA-A6DE-01	M0
TCGA-BA-A6DF-01	M0
TCGA-BA-A6DF-01	M0
TCGA-BA-A6DL-01	M0
TCGA-BA-A8YP-01	M0
TCGA-BA-A8YP-01	M0
TCGA-BB-7864-01	M0
TCGA-BB-7866-01	M0
TCGA-BB-7866-01	M0
TCGA-C9-A480-01	M0
TCGA-CN-4722-01	M0
TCGA-CN-4722-01	M0

TCGA-CN-4734-01	M0
TCGA-CN-4734-01	M0
TCGA-CN-5361-01	M0
TCGA-CN-5361-01	M0
TCGA-CQ-6219-01	R
TCGA-CQ-6219-01	R
TCGA-CQ-6222-01	M0
TCGA-CQ-6222-01	M0
TCGA-CQ-7063-01	M0
TCGA-CQ-7064-01	M0
TCGA-CQ-7064-01	M0
TCGA-CQ-A4C9-01	M0
TCGA-CQ-A4CA-01	M0
TCGA-CQ-A4CA-01	M0
TCGA-CR-5249-01	M0
TCGA-CR-5250-01	M0
TCGA-CR-5250-01	M0
TCGA-CV-5430-01	M0
TCGA-CV-5431-01	M0
TCGA-CV-5431-01	M0
TCGA-CV-7427-01	M0
TCGA-CV-7428-01	M0
TCGA-CV-7428-01	M0
TCGA-CV-7446-01	M0
TCGA-CV-7446-01	M0
TCGA-CV-7568-01	M0
TCGA-CV-7568-01	M0

TCGA-F7-7848-01	M0
TCGA-F7-8298-01	M0
TCGA-F7-8298-01	M0
TCGA-H7-8501-01	M0
TCGA-H7-8502-01	M0
TCGA-H7-8502-01	M0
TCGA-H7-A6C4-01	M0
TCGA-H7-A6C5-01	M0
TCGA-H7-A6C5-01	M0
TCGA-IQ-A61J-01	M0
TCGA-IQ-A61K-01	M0
TCGA-IQ-A61K-01	M0
TCGA-IQ-A61L-01	M0
TCGA-IQ-A61L-01	M0
TCGA-KU-A6H7-06	M0
TCGA-KU-A6H7-06	M0
TCGA-QK-A6VC-01	M0
TCGA-QK-A8Z7-01	M0
TCGA-QK-A8Z7-01	M1
TCGA-QK-A8Z8-01	M1
TCGA-QK-A8Z8-01	R
TCGA-QK-A8Z9-01	R
TCGA-QK-A8Z9-01	R
TCGA-QK-A8ZA-01	M0
TCGA-QK-A8ZB-0	M0
TCGA-QK-A8ZB-01	M0
TCGA-QK-AA3J-01	M0

TCGA-QK-AA3J-01	M0
TCGA-UF-A71A-06	M0
TCGA-UF-A71A-06	M0
OT10_N	N
OT13_N	N
OT15_N	N
OT16_N	N
OT18_N	N
OT20_N	N
OT21_N	N
OT23_N	N
OT25_N	N
OT27_N	N
OT29_N	N
OT30_N	N
OT31_N	N
OT32_N	N
OT33_N	N
OT34_N	N
OT35_N	N
OT36_N	N
OT37_N	N
OT38_N	N
OT39_N	N
OT4_N	N
OT40_N	N
OT41_N	N

OT42_N	N
OT43_N	N
OT45_N	N
OT46_N	N
OT47_N	N
OT48_N	N
OT49_N	N
OT5_N	N
OT50_N	N
OT53_N	N
OT55_N	N
OT7_N	N
OT9_N	N

Supplementary Table S3: Oral tongue squamous cell carcinoma (OTSCC) with all four events assayed within the same tumor, and their clinical attributes used for validation.

Sample ID	R (Loco-regional recurrence); M1(metastasis); M0 (Non-metastatic/Non-recurrent) tumors; N (Solid Tissue Normals).
OT1_T	M0
OT11_T	M0
OT12_T	M0
OT14_T	M0

OT17_T	M0
OT19_T	M0
OT2_T	M1
OT22_T	M0
OT24_T	M0
OT26_T	M0
OT28_T	M0
OT52_T	M0
OT54_T	M0
OT3_T	M0
OT6_T	M0
OT8_T	M0
OT44_T	R
OT51_T	M1
OT1_N	N
OT11_N	N
OT12_N	N
OT14_N	N
OT17_N	N
OT19_N	N
OT2_N	N
OT22_N	N
OT24_N	N
OT26_N	N
OT28_N	N
OT52_N	N

OT54_N	N
OT3_N	N
OT6_N	N
OT8_N	N
OT44_N	N
OT51_N	N

Supplementary Table S4: Discovered events associated with Metastases and Recurrence in TCGA HNSCC, pathways mapped, confirmation and validation status.

user_selected_clinical_variable	event_affected_gene	event_type	event	sample_frequency	event_priority	filter_type	cancer_pathways	confirmation	validation	minimal_discovery_panel
Met/REC	ACSS3	cnv	D_ACSS3	1.20482	D	event_priority_filter	NA	YES	YES	YES
Met/REC	AGO3	cnv	D_AGO3	1.20482	D	event_priority_filter	NA	YES	YES	YES
Met/REC	ANAPC4	cnv	D_ANAPC4	1.20482	D	event_priority_filter	cell_cycle	YES	YES	NO
Met/REC	ANKFY1	fcnv	D_ANKFY1	2.40964	B	event_priority_filter	NA	YES	YES	NO
Met/REC	ANO5	cnv	I_ANO5	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	ANXA11	fcnv	I_ANXA11	1.20482	B	event_priority_filter	NA	YES	YES	YES
Met/REC	ARHGAP10	cnv	I_ARHGAP10	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	ARHGEF38	cnv	D_ARHGEF38	1.20482	D	event_priority_filter	NA	YES	YES	YES
Met/REC	CALY	cnv	D_CALY	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	CCDC97	fcnv	I_CCDC97	1.20482	B	event_priority_filter	NA	YES	YES	NO
Met/REC	CCKAR	cnv	D_CCKAR	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	CCPG1	fcnv	I_CCPG1	1.20482	B	event_priority_filter	NA	YES	YES	NO
Met/REC	CIC	fcnv	I_CIC	1.20482	B	event_priority_filter	NA	YES	YES	NO
Met/REC	CNDP1	cnv	I_CNDP1	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	CNKSR3	cnv	D_CNKSR3	2.40964	D	event_priority_filter	NA	YES	YES	NO

Met/REC	COPS2	fcnv	I_COPS2	1.20482	B	event_priority_filter	NA	YES	YES	NO
Met/REC	CSRP3	cnv	I_CSRP3	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	DBX1	cnv	I_DBX1	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	DCLK2	cnv	I_DCLK2	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	DOCK2	cnv	D_DOCK2	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	DVL2	fcnv	D_DVL2	1.20482	B	event_priority_filter	mtor_wnt	YES	YES	YES
Met/REC	ECHS1	cnv	D_ECHS1	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	ERF	fcnv	I_ERF	1.20482	B	event_priority_filter	NA	YES	YES	NO
Met/REC	FBXW7	cnv	I_FBXW7	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	FGF10	cnv	D_FGF10	1.20482	D	event_priority_filter	mapk_pi3k_akt	YES	YES	YES
Met/REC	FGFR4	fcnv	I_FGFR4	1.20482	B	event_priority_filter	mapk_pi3k_akt	YES	YES	YES
Met/REC	FRMPD4	cnv	I_FRMPD4	2.40964	D	event_priority_filter	NA	YES	YES	YES
Met/REC	FUOM	cnv	D_FUOM	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	GAS2	cnv	I_GAS2	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	GBE1	cnv	I_GBE1	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	GCSAML	cnv	D_GCSAML	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	GNPDA2	cnv	D_GNPDA2	3.61446	D	event_priority_filter	NA	YES	YES	NO
Met/REC	GRXCR1	cnv	D_GRXCR1	3.61446	D	event_priority_filter	NA	YES	YES	NO

Met/REC	GTF2H1	cnv	I_GTF2H1	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	GUF1	cnv	D_GUF1	3.61446	D	event_priority_filter	NA	YES	YES	NO
Met/REC	HK2	cnv	D_HK2	1.20482	D	event_priority_filter	NA	YES	YES	YES
Met/REC	HTATIP2	cnv	I_HTATIP2	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	IGSF22	cnv	I_IGSF22	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	KCNC1	cnv	I_KCNC1	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	LDHAL6A	cnv	I_LDHAL6A	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	MAGI3	cnv	I_MAGI3	1.20482	D	event_priority_filter	NA	YES	YES	YES
Met/REC	MAT1A	fcnv	I_MAT1A	1.20482	B	event_priority_filter	NA	YES	YES	NO
Met/REC	MIR573	cnv	I_MIR573	1.20482	D	event_priority_filter	NA	YES	YES	YES
Met/REC	MMAA	cnv	I_MMAA	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	MORC3	fcnv	I_MORC3	1.20482	B	event_priority_filter	NA	NO	NA	NO
Met/REC	MORF4L1	fcnv	D_MORF4L1	1.20482	B	event_priority_filter	NA	YES	YES	YES
Met/REC	MPHOSPH6	cnv	D_MPHOSPH6	1.20482	D	event_priority_filter	NA	YES	YES	YES
Met/REC	MRGPRX2	cnv	I_MRGPRX2	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	MTHFS	fcnv	D_MTHFS	1.20482	B	event_priority_filter	NA	YES	YES	YES
Met/REC	MUC15	cnv	I_MUC15	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	MYO5C	fcnv	I_MYO5C	1.20482	B	event_priority_filter	NA	YES	YES	NO

Met/REC	NAV2	cnv	I_NAV2	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	NEDD4	fcnv	I_NEDD4	1.20482	B	event_priority_filter	NA	YES	YES	NO
Met/REC	NELL1	cnv	I_NELL1	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	NLRP3	cnv	D_NLRP3	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	NNT	cnv	D_NNT	1.20482	D	event_priority_filter	NA	YES	YES	YES
Met/REC	NR3C2	cnv	I_NR3C2	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	NSUN5	fcnv	D_NSUN5	1.20482	B	event_priority_filter	NA	YES	YES	NO
Met/REC	OR2B11	cnv	D_OR2B11	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	OR2C3	cnv	D_OR2C3	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	OXCT1	cnv	D_OXCT1	1.20482	D	event_priority_filter	NA	YES	YES	YES
Met/REC	PFKM	fcnv	D_PFKM	1.20482	B	event_priority_filter	NA	YES	YES	NO
Met/REC	PLAUR	fcnv	I_PLAUR	1.20482	B	event_priority_filter	NA	YES	YES	YES
Met/REC	PRAP1	cnv	D_PRAP1	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	PRMT3	cnv	I_PRMT3	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	PTPN5	cnv	I_PTPN5	2.40964	D	event_priority_filter	mapk	YES	YES	NO
Met/REC	PTPRR	cnv	D_PTPRR	1.20482	D	event_priority_filter	mapk	NO	NA	NO
Met/REC	RAB27A	fcnv	I_RAB27A	1.20482	B	event_priority_filter	NA	YES	YES	NO
Met/REC	RERGL	cnv	D_RERGL	1.20482	D	event_priority_filter	NA	YES	YES	YES

Met/REC	RFC2	fcnv	D_RFC2	1.20482	B	event_priority_filter	NA	YES	YES	NO
Met/REC	RICTOR	cnv	D_RICTOR	1.20482	D	event_priority_filter	mtor	YES	YES	YES
Met/REC	RPAP2	cnv	D_RPAP2	1.20482	D	event_priority_filter	NA	YES	YES	YES
Met/REC	RPL26L1	fcnv	I_RPL26L1	1.20482	B	event_priority_filter	NA	YES	YES	YES
Met/REC	RPS3A	cnv	I_RPS3A	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	SAA4	cnv	I_SAA4	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	SAAL1	cnv	I_SAAL1	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	SEL1L3	cnv	D_SEL1L3	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	SENP1	fcnv	D_SENP1	1.20482	B	event_priority_filter	NA	YES	YES	NO
Met/REC	SERGEF	cnv	I_SERGEF	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	SH3D19	cnv	I_SH3D19	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	SLC17A6	cnv	I_SLC17A6	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	SLC34A2	cnv	D_SLC34A2	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	SLC5A12	cnv	I_SLC5A12	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	SMAD1	cnv	I_SMAD1	1.20482	D	event_priority_filter	tgf	YES	YES	NO
Met/REC	SMIM20	cnv	D_SMIM20	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	SMIM21	cnv	I_SMIM21	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	SPTY2D1	cnv	I_SPTY2D1	2.40964	D	event_priority_filter	NA	YES	YES	NO

Met/REC	SRBD1	cnv	D_SRBD1	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	SSBP2	cnv	D_SSBP2	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	SVIP	cnv	I_SVIP	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	SYDE2	cnv	D_SYDE2	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	TBC1D19	cnv	D_TBC1D19	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	TMEM154	cnv	I_TMEM154	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	TMEM155	fcnv	I_TMEM155	1.20482	B	event_priority_filter	NA	YES	YES	NO
Met/REC	TMEM86A	cnv	I_TMEM86A	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	TPH1	cnv	I_TPH1	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	TSG101	cnv	I_TSG101	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	TSHZ1	cnv	I_TSHZ1	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	UEVLD	cnv	I_UEVLD	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	USP32	cnv	D_USP32	1.20482	D	event_priority_filter	NA	YES	YES	YES
Met/REC	USP8	fcnv	I_USP8	1.20482	B	event_priority_filter	NA	YES	YES	NO
Met/REC	YIPF7	cnv	D_YIPF7	3.61446	D	event_priority_filter	NA	YES	YES	NO
Met/REC	ZADH2	cnv	I_ZADH2	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	ZBTB7A	fcnv	I_ZBTB7A	1.20482	B	event_priority_filter	NA	YES	YES	YES
Met/REC	ZCCHC4	cnv	D_ZCCHC4	1.20482	D	event_priority_filter	NA	YES	YES	NO

Met/REC	ZDHHC13	cnv	I_ZDHHC13	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	ZNF407	cnv	I_ZNF407	2.40964	D	event_priority_filter	NA	YES	YES	NO
Met/REC	ZNF511	cnv	D_ZNF511	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	ZNF827	cnv	I_ZNF827	1.20482	D	event_priority_filter	NA	YES	YES	NO
Met/REC	ACAD8	mut	ACAD8	3.61446	A	event_priority_filter	NA	NO	NA	NO
Met/REC	AIMP1	expr	AIMP1	7.22892	NA	sample_frequency_filter	NA	YES	YES	YES
Met/REC	C1QA	expr	C1QA	4.81928	E	event_priority_filter	NA	NO	NA	NO
Met/REC	C1QB	expr	C1QB	4.81928	E	event_priority_filter	NA	NO	NA	NO
Met/REC	CHPF2	expr	CHPF2	6.0241	E	event_priority_filter	NA	NO	NA	NO
Met/REC	CHPF2	expr	CHPF2	6.0241	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	CIC	expr	CIC	7.22892	E	event_priority_filter	NA	YES	YES	YES
Met/REC	CIC	expr	CIC	7.22892	NA	sample_frequency_filter	NA	YES	YES	YES
Met/REC	CTU1	expr	CTU1	4.81928	E	event_priority_filter	NA	YES	YES	YES
Met/REC	CTU1	expr	CTU1	4.81928	NA	sample_frequency_filter	NA	YES	YES	YES
Met/REC	DOHH	expr	DOHH	7.22892	NA	sample_frequency_filter	NA	YES	YES	YES
Met/REC	DQX1	expr	DQX1	4.81928	E	event_priority_filter	NA	NO	NA	NO
Met/REC	DQX1	expr	DQX1	4.81928	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	ECE2	expr	ECE2	6.0241	E	event_priority_filter	NA	NO	NA	NO

Met/REC	ECE2	expr	ECE2	6.0241	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	EEF1B2	mut	EEF1B2	1.20482	A	event_priority_filter	NA	YES	YES	NO
Met/REC	FAM102A	expr	FAM102A	4.81928	E	event_priority_filter	NA	NO	NA	NO
Met/REC	FAM102A	expr	FAM102A	4.81928	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	FAM134C	expr	FAM134C	6.0241	E	event_priority_filter	NA	YES	YES	YES
Met/REC	FAM134C	expr	FAM134C	6.0241	NA	sample_frequency_filter	NA	YES	YES	YES
Met/REC	FOXO4	expr	FOXO4	6.0241	E	event_priority_filter	NA	NO	NA	NO
Met/REC	GMFB	expr	GMFB	6.0241	E	event_priority_filter	NA	YES	YES	YES
Met/REC	GTF2IRD1	expr	GTF2IRD1	4.81928	E	event_priority_filter	NA	NO	NA	NO
Met/REC	ITGA9	mut	ITGA9	1.20482	A	event_priority_filter	ecm_focal_adhesion_pi3k_akt	YES	YES	YES
Met/REC	KLF4	mut	KLF4	1.20482	A	event_priority_filter	NA	YES	YES	NO
Met/REC	KLHDC7B	expr	KLHDC7B	6.0241	E	event_priority_filter	NA	NO	NA	NO
Met/REC	KLHDC7B	expr	KLHDC7B	6.0241	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	LIMS1	expr	LIMS1	6.0241	E	event_priority_filter	NA	YES	YES	YES
Met/REC	LIMS1	expr	LIMS1	6.0241	NA	sample_frequency_filter	NA	YES	YES	YES
Met/REC	MORC2	expr	MORC2	4.81928	E	event_priority_filter	NA	NO	NA	NO
Met/REC	MORC2	expr	MORC2	4.81928	NA	sample_frequency_filter	NA	NO	NA	NO

Met/REC	MRS2	expr	MRS2	7.22892	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	NPM3	expr	NPM3	7.22892	E	event_priority_filter	NA	NO	NA	NO
Met/REC	NPM3	expr	NPM3	7.22892	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	NRAS	mut	NRAS	1.20482	A	event_priority_filter	apoptosis_mapk_mtor_pi3k_akt_vegf	YES	YES	YES
Met/REC	PCBP1	mut	PCBP1	2.40964	A	event_priority_filter	NA	YES	YES	NO
Met/REC	PGS1	expr	PGS1	4.81928	E	event_priority_filter	NA	NO	NA	NO
Met/REC	PGS1	expr	PGS1	4.81928	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	PJA2	expr	PJA2	8.43373	E	event_priority_filter	NA	NO	NA	NO
Met/REC	PJA2	expr	PJA2	8.43373	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	PKP4	expr	PKP4	7.22892	E	event_priority_filter	NA	NO	NA	NO
Met/REC	PKP4	expr	PKP4	7.22892	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	PPIAL4C	expr	PPIAL4C	7.22892	NA	sample_frequency_filter	NA	YES	YES	YES
Met/REC	PPIC	expr	PPIC	4.81928	E	event_priority_filter	NA	NO	NA	NO
Met/REC	REXO2	expr	REXO2	6.0241	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	RNF139	expr	RNF139	7.22892	E	event_priority_filter	NA	NO	NA	NO
Met/REC	RNF139	expr	RNF139	7.22892	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	SFRP1	expr	SFRP1	4.81928	NA	sample_frequency_filter	wnt	NO	NA	NO

Met/REC	SIGLEC1	expr	SIGLEC1	3.61446	E	event_priority_filter	NA	NO	NA	NO
Met/REC	SLC9A5	expr	SLC9A5	4.81928	E	event_priority_filter	NA	NO	NA	NO
Met/REC	SLC9A5	expr	SLC9A5	4.81928	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	TFDP1	mut	TFDP1	2.40964	A	event_priority_filter	cell_cycle_tgf	NO	NA	NO
Met/REC	ZBP1	expr	ZBP1	3.61446	E	event_priority_filter	NA	NO	NA	NO
Met/REC	ACAT2	meth	ACAT2_cg23500094	7.22892	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	AGAP1	meth	AGAP1_cg08805497	6.0241	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	ALDOA	fmeth	ALDOA_ALDOA_cg07025134	2.40964	C	event_priority_filter	NA	NO	NA	NO
Met/REC	AMH	meth	AMH_cg26659805	7.22892	NA	sample_frequency_filter	camp_cytokine_tgf	NO	NA	NO
Met/REC	ASPSCR1	meth	ASPSCR1_cg06238300	6.0241	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	BLNK	fmeth	BLNK_BLNK_cg02980499	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	BRMS1	meth	BRMS1_cg21582163	7.22892	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	CAV1	fmeth	CAV1_MIR107_cg24445034	1.20482	C	event_priority_filter	focal_adhesion	YES	YES	YES
Met/REC	CCDC59	fmeth	CCDC59_CCDC59_cg01903903	1.20482	C	event_priority_filter	NA	YES	YES	NO
Met/REC	CCND2	meth	CCND2_cg03801902	6.0241	NA	sample_frequency_filter	cell_cycle_focal_adhesion_jak_stat_p53_pi3k_akt_wnt	NO	NA	NO
Met/REC	CHAF1A	meth	CHAF1A_cg23811775	9.63855	NA	sample_frequency_filter	NA	NO	NA	NO

Met/REC	CHST12	meth	CHST12_cg10145585	6.0241	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	CRABP2	meth	CRABP2_cg27493997	6.0241	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	CROCC	fmeth	CROCC_CROCC_cg23732962	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	CYBA	fmeth	CYBA_CYBA_cg24046411	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	CYBA	fmeth	CYBA_CYBA_cg27176614	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	DCAF12L1	meth	DCAF12L1_cg01913196	4.81928	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	DLG2	meth	DLG2_cg10901805	7.22892	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	DPP9	meth	DPP9_cg24423897	7.22892	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	DPYSL2	meth	DPYSL2_cg02835343	7.22892	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	EGLN1	meth	EGLN1_cg20015535	7.22892	NA	sample_frequency_filter	NA	YES	YES	YES
Met/REC	EIF5A	meth	EIF5A_cg24699433	7.22892	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	EIF5A2	fmeth	EIF5A2_EIF5A2_cg12598803	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	EIF5A2	fmeth	EIF5A2_EIF5A2_cg26209058	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	EN1	meth	EN1_cg02739346	4.81928	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	EP400NL	fmeth	EP400NL_EP400NL_cg25632983	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	ERBB2	fmeth	ERBB2_MIR125A_cg19782652	1.20482	C	event_priority_filter	adherens_focal_adhesion	YES	YES	NO
Met/REC	ERBB2	fmeth	ERBB2_MIR134_cg26238975	1.20482	C	event_priority_filter	adherens_focal_adhes	NO	NA	NO

							ion			
Met/REC	FBLN5	meth	FBLN5_cg21883802	7.22892	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	FBN1	meth	FBN1_cg12975862	4.81928	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	FBXO43	meth	FBXO43_cg23700125	9.63855	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	FNTA	fmeth	FNTA_FNTA_cg07294234	2.40964	C	event_priority_filter	NA	YES	YES	YES
Met/REC	GALNT2	meth	GALNT2_cg00589617	6.0241	F	event_priority_filter	NA	NO	NA	NO
Met/REC	GALNT2	meth	GALNT2_cg00589617	6.0241	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	GDAP1	fmeth	GDAP1_GDAP1_cg22105613	2.40964	C	event_priority_filter	NA	NO	NA	NO
Met/REC	GLRX5	meth	GLRX5_cg26007049	7.22892	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	GNB4	fmeth	GNB4_GNB4_cg01750654	1.20482	C	event_priority_filter	pi3k_akt	NO	NA	NO
Met/REC	HAUS5	fmeth	HAUS5_HAUS5_cg05296832	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	HEATR1	meth	HEATR1_cg00907549	7.22892	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	HNRNPM	fmeth	HNRNPM_HNRNPM_cg05923062	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	HOMER3	fmeth	HOMER3_HOMER3_cg19788957	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	HTR3E	fmeth	HTR3E_HTR3E_cg11158819	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	KLHL20	meth	KLHL20_cg11720170	7.22892	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	LBH	meth	LBH_cg25072436	7.22892	NA	sample_frequency_filter	NA	NO	NA	NO

Met/REC	LMX1A	fmeth	LMX1A_LMX1A_cg23971170	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	LRRK1	meth	LRRK1_cg24163563	6.0241	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	MAGEA10	fmeth	MAGEA10_MAGEA10_cg19964192	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	MAN2B1	meth	MAN2B1_cg02484047	8.43373	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	MAPK12	meth	MAPK12_cg21974239	8.43373	NA	sample_frequency_filter	mapk_vegf	NO	NA	NO
Met/REC	MBP	meth	MBP_cg00989538	6.0241	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	MED11	meth	MED11_cg09832947	8.43373	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	MORF4L1	fmeth	MORF4L1_MORF4L1_cg20905746	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	NLGN2	meth	NLGN2_cg10092265	7.22892	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	NOS3	meth	NOS3_cg12547085	7.22892	NA	sample_frequency_filter	pi3k_akt_vegf	NO	NA	NO
Met/REC	NR1H2	meth	NR1H2_cg14673743	4.81928	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	PCCB	meth	PCCB_cg05582311	7.22892	NA	sample_frequency_filter	NA	YES	YES	YES
Met/REC	PCNT	fmeth	PCNT_PCNT_cg13822122	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	PNMT	fmeth	PNMT_PNMT_cg16268778	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	PPP1R1B	fmeth	PPP1R1B_PPP1R1B_cg06068801	1.20482	C	event_priority_filter	camp	NO	NA	NO
Met/REC	PPP1R1B	fmeth	PPP1R1B_PPP1R1B_cg09762778	1.20482	C	event_priority_filter	camp	NO	NA	NO

Met/REC	PRF1	fmeth	PRF1_PRF1_cg12336290	1.20482	C	event_priority_filter	apoptosis	NO	NA	NO
Met/REC	RBMS2	meth	RBMS2_cg23002907	8.43373	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	RNF138	meth	RNF138_cg22141237	4.81928	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	RPLP2	meth	RPLP2_cg01485797	6.0241	NA	sample_frequency_filter	NA	YES	YES	YES
Met/REC	RPRD1A	fmeth	RPRD1A_RPRD1A_cg00019809	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	RPRD1A	fmeth	RPRD1A_RPRD1A_cg19629710	1.20482	C	event_priority_filter	NA	YES	YES	NO
Met/REC	RYBP	meth	RYBP_cg04421631	7.22892	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	SLC45A4	meth	SLC45A4_cg08829393	7.22892	NA	sample_frequency_filter	NA	YES	YES	YES
Met/REC	SNUPN	meth	SNUPN_cg22258732	4.81928	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	SORBS3	meth	SORBS3_cg07924703	6.0241	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	SPIRE2	meth	SPIRE2_cg07402062	8.43373	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	SPOPL	fmeth	SPOPL_SPOPL_cg09464263	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	STIM1	meth	STIM1_cg04346968	6.0241	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	TAF1D	fmeth	TAF1D_TAF1D_cg10353388	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	TEX11	meth	TEX11_cg08167097	7.22892	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	TNK2	meth	TNK2_cg03207310	4.81928	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	TNPO1	meth	TNPO1_cg05579703	6.0241	NA	sample_frequency_filter	NA	NO	NA	NO

Met/REC	TOLLIP	meth	TOLLIP_cg01369233	6.0241	NA	sample_frequency_filter	NA	YES	YES	YES
Met/REC	TRRAP	meth	TRRAP_cg03457229	6.0241	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	UMPS	meth	UMPS_cg15050328	6.0241	NA	sample_frequency_filter	NA	YES	YES	YES
Met/REC	UST	meth	UST_cg21273029	6.0241	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	WDR66	meth	WDR66_cg01765077	6.0241	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	ZFHX3	meth	ZFHX3_cg05704496	6.0241	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	ZIC3	fmeth	ZIC3_ZIC3_cg10451760	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	ZNF121	meth	ZNF121_cg06998238	6.0241	NA	sample_frequency_filter	NA	YES	YES	YES
Met/REC	ZNF410	fmeth	ZNF410_ZNF410_cg12152256	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	ZNF490	meth	ZNF490_cg24093429	6.0241	NA	sample_frequency_filter	NA	YES	YES	YES
Met/REC	ZNF497	fmeth	ZNF497_ZNF497_cg14989522	1.20482	C	event_priority_filter	NA	NO	NA	NO
Met/REC	ZNF516	meth	ZNF516_cg16275118	9.63855	F	event_priority_filter	NA	NO	NA	NO
Met/REC	ZNF516	meth	ZNF516_cg16275118	9.63855	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	ZNF580	meth	ZNF580_cg26209676	4.81928	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	ZNF581	meth	ZNF581_cg26209676	4.81928	NA	sample_frequency_filter	NA	NO	NA	NO
Met/REC	ZNF791	meth	ZNF791_cg24093429	6.0241	NA	sample_frequency_filter	NA	YES	YES	YES