RESEARCH ARTICLE

Editorial Process: Submission:06/06/2025 Acceptance:02/26/2026 Published:03/06/2026

Meta-Merging the Transcriptomes of Gastric Tumors Redefines the Connections among Molecular and Clinical Subtypes

Balaji T Sekar, Kumaresan Ganesan*

Abstract

Introduction: The availability of a large number of cancer expression profiles presents an excellent opportunity to re-investigate various biological and clinical questions. While several expression profiles have been established for different cancers, merging them may provide a more powerful platform for extensively extrapolating molecular and clinical features across multiple cohorts. **Materials and Methods:** In this study, five gastric tumor expression profiles from the Gene Expression Omnibus [GEO] and one in-house cohort comprising a total of 1,060 samples were merged. The batch effect was removed using non-parametric ComBat analysis, and the seamless merging of datasets was confirmed through various parameters. **Results:** Extrapolation of ACRG [Asian Cancer Research Group] and TCGA [The Cancer Genome Atlas] molecular subtypes in the merged cohort of 1,060 gastric tumors revealed nine distinct clusters. Notably, the following patterns were observed: [i] mutual exclusivity between Epithelial to Mesenchymal Transition [EMT] and Microsatellite Instability [MSI] subtypes in 90% of tumors; [ii] overlapping occurrence of EMT and MSI subtypes in the remaining tumors; [iii] overlap between MSI and Epstein-Barr Virus [EBV] subtype tumors; [iv] both commonalities and differences between EMT and Genomically Stable [GS] subtypes; and [v] an association between EBV positivity and PI3K mutation. **Conclusion:** The current study demonstrates that compiling a larger expression profile is valuable for revisiting the molecular features and epidemiology associated with molecular subtypes, thereby aiding in the development of novel diagnostics and targeted therapeutics.

Keywords: Meta-merge- gastric tumors- batch QC- EMT

Asian Pac J Cancer Prev, **27 (3)**, 883-895

Introduction

Being the fifth leading cancer in the world, the third leading cause of cancer-related death, and responsible for almost 800,000 deaths every year, gastric cancer remains a global issue. About 70% of the incidence of gastric cancer occurs in developing countries [1]. The availability of a large quantum of genomics data is resourceful for researchers to explore developing the leads for precision medicine [2]. The Gene Expression Omnibus [GEO] database has genome-wide profiles of over 2 million samples. Exploring multiple cohorts of expression profiles from various studies is expected to enhance the reliability of observations, statistical power, and reduce the lab-specific effects [3, 4]. However, integrating the gene expression profiles from different experiments is challenging [5]. The major expression profiling platforms include Affymetrix, Illumina, Agilent, and RNA sequencing. The datasets from different platforms provide an opportunity for intra- and inter-dataset comparisons. However, there are a few issues associated with meta-analysis and merging of gene expression profiles [6]. The different platforms introduce distinct technological biases and produce data with different shapes and scales [7]. The pipelines for background correction and normalization of array-based gene expression include Robust Multiarray Average [RMA], frozen Robust Multiarray Analysis [fRMA], Single Channel Array Normalization [SCAN], and Universal exPression Code [UPC] [2].

Various methods used for merging microarray datasets include SVA, ComBat, RMA, RMA and ComBat, and scaling [8]. Expression data of idiopathic pulmonary fibrosis [IPF] and sarcoidosis were merged to find the genes discriminating against IPF and sarcoidosis [9]. Combining data from different diseases using common clinical manifestations was also explored [10]. Merging multiple datasets was employed to define a novel signature for the prediction of relapse-free survival in colon cancer [11]. Merging ten datasets of non-small cell lung tumors was found useful for exploratory genomes [2]. In the current study, six gastric cancer expression profiles were merged and explored.

The batch effect is the major hurdle in integrating data from multiple expression profiles [12]. Across different

*Unit of Excellence in Cancer Genetics, Department of Genetics, Centre for Excellence in Genomic Sciences, School of Biological Sciences, Madurai Kamaraj University, Madurai, India. *For Correspondence: kumaresan@mkuniversity.ac.in*

expression profiles, the variation stem from technical factors such as differences in profiling platforms, lab protocols, experimenter, processing procedures, reagent batch, and the statistical methods employed [13]. Batch effect is defined as i] the uncontrollable errors unrelated to the biological variation [14], ii] the cumulative errors introduced by time and place-dependent experimental variations [15], iii] sub-groups of measurements that are unrelated to the biological variables [16], iv] systematic differences between the measurements of different batches of experiments [17], and v] systematic technical differences in different batches [18]. To sort out the batch effect, Batch QC Bioconductor R package was developed [13]. Similarly, the sva package was introduced for identifying and removing batch effects [19]. The parametric and non-parametric empirical Bayes framework method was used for adjusting batch effects [3]. Non-parametric ComBat analysis was established for combining genomic data [3, 5]. Modified-ComBat analysis was developed to enable gene expression meta-analyses to combine data from different studies [20]. ComBat analysis is the widely used method for the batch effect removal [12]. A combination of ComBat analysis and quantile normalization has been reported to be the best approach for the removal of batch effects in longitudinal gene expression data [21].

Gastric cancer also exhibits heterogeneity among patients [22, 23]. Molecular classification of gastric cancer and targeted therapeutics are believed to help in improving the outcome of patients [24]. The Cancer Genome Atlas [TCGA] and Asia Cancer Research Group [ACRG] studies have defined molecular subtypes [25, 26]. Subtype-specific systemic therapy is expected to provide better efficacy in treatment strategies [24].

In the current study, ACRG and TCGA subtype-specific gene sets were explored for their clinical relevance in gastric cancer by using meta-merged cohort 1060. With the large number of samples and their clinical information, the expression patterns of molecular subtypes were analyzed. Based on the expression pattern, clusters were separated, and cluster-specific clinical features were inferred. Notably, i] exclusivity between EMT and MSI subtypes in 90 % of tumors, ii] an overlap between the occurrence of MSI and EMT subtypes in 10 % of tumors, and iii] an association between EBV and PI3K mutations, iv] association between EBV and MSI tumors, and v} possible stratification of GS gastric tumors as GS-EMT and GS-non EMT were inferred upon extrapolating the clinical features across cohorts.

Materials and Methods

Genome-wide expression profiles

Whole genome expression profiles of gastric tumors were collected from Gene Expression Omnibus [GEO]. The collected six expression profiles of gastric tumors include GSE15459 [n=200], GSE54129 [n=111], GSE62254 [n=300], GSE29272 [n=134], and TCGA [n=265]. In addition, we also included the in-house genome-wide expression profile of gastric tumors GSE146996 [n=50]. The details of the profiles included in the study are provided in Figure 1a.

Pre-process of expression profiles

Merging of multiple cohorts requires pre-processing of the individual expression profile datasets. Individual datasets used for the analysis were normalized by Robust Multi-Array Average [RMA]. Raw .CEL files collected from GEO were RMA normalized using the affy package. Expression values of multiple probes of the single gene were averaged using dChip [27].

Merging datasets

All the datasets were arranged in individual worksheets of Microsoft Excel. Common genes present in all the datasets were short-listed. With the 'vlookup' function in Excel, the expression values of individual datasets were merged one by one for all the genes common to the datasets used in the study. All datasets were merged into one cohort as meta-merged cohort-1060. Meta-merge cohort-1060 contains gastric tumor samples from the profiles GSE15459, GSE54129, GSE62254, GSE29272, TCGA, and GSE146996. The merged cohort was analyzed for its batch effect, and the Batch QC [13] package was used to remove the batch effect.

Gene set based pathway activation scoring analysis

Gene signatures of gastric cancer molecular subtypes of ACRG and TCGA studies were used for the analysis. Z score-based pathway activation scoring was performed for both ACRG and TCGA subtype gene signatures in the meta-merged cohort-1060 with Perl and R scripts [26, 28]. dChip tool was used for the hierarchical clustering of gene sets in the merged cohort. Based on the clustering of expression pattern visualized in the heat map, nine clusters were identified and analyzed for their clinical and molecular features.

Results

Meta-merging of six gastric tumor cohorts

The merging of multiple cohorts of the expression profiles of tumors would pave ways to extrapolate the clinical, diagnostic and therapeutic information from the well-established cohorts to other cohorts. The available and selected expression profiles of gastric tumors were collected from Gene Expression Omnibus [GEO]. Five expression profiles of gastric tumors, GSE15459 [n=200], GSE54129 [n=111], GSE62254 [n=300], GSE29272 [n=134], and TCGA [n=265] from GEO were used for the investigation. In addition, an in-house genome-wide expression profile of gastric tumors from GSE146996 [n=50] was used. The details of the six gastric cancer profiles used in the study, the platform used, and their details are shown in Figure 1a.

To merge multiple cohorts, pre-processing of the expression profiles was performed through the following steps: i] uniform normalization of all the profiles by Robust Multi-Array Average [RMA] normalization, ii] merging of duplicate probes of a gene, iii] arranging the datasets according to their total number of probes, and iv] matching the genes across datasets and merging the profiles. The detailed steps involved in merging the datasets are described in methods. The meta-merged cohort comprises

a

	Cohort	Samples	Year	Country
META-MERGE (1060)	GSE15459	200	2009	Singapore
	GSE62254	300	2015	Korea
	GSE54129	111	2017	China
	GSE146996	50	2020	Madurai
	TCGA	265	2009	Multi-national
	GSE29272	134	2011	USA

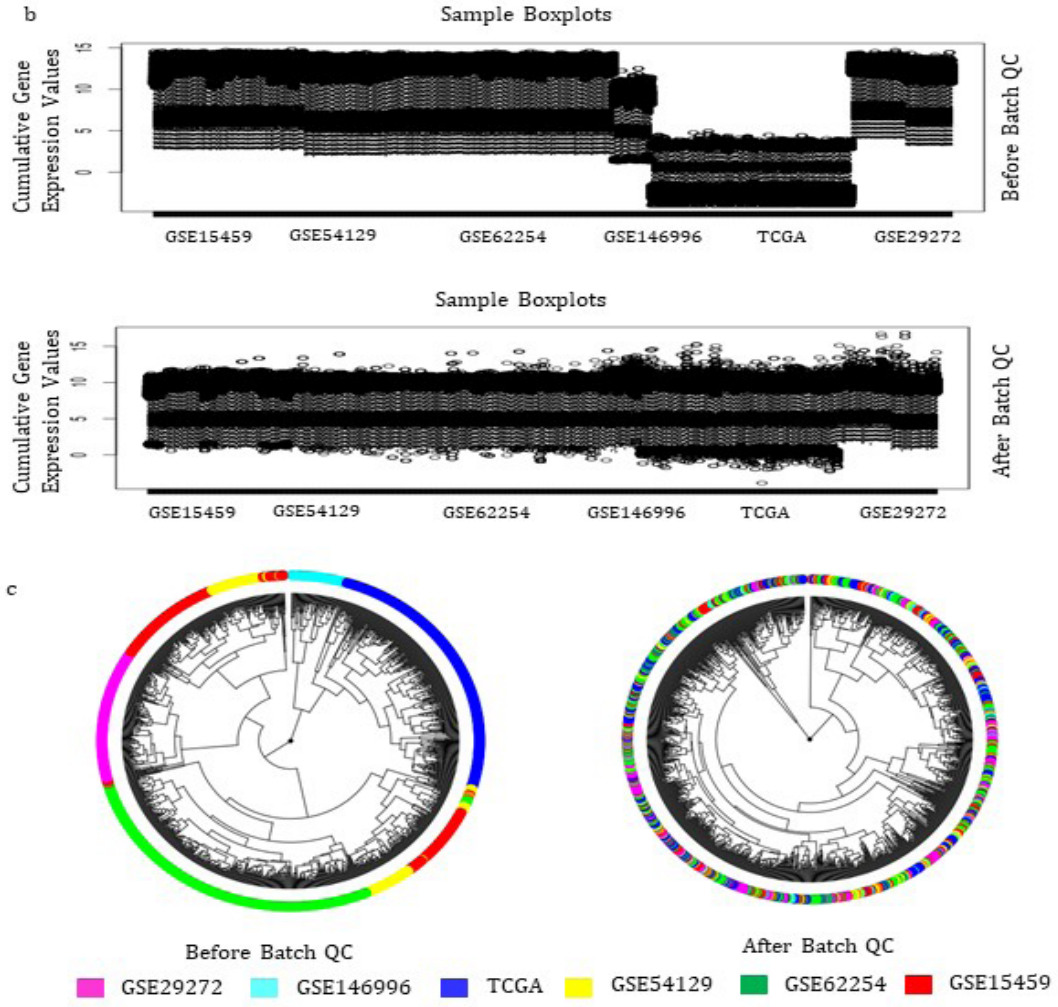


Figure 1. Seamless Merging of 6 Expression Profile Datasets to a Meta-Merge Cohort-1060. [a] Details of the gastric tumor profiles used in the study. [b] Boxplots represent the expression values of samples in six gastric cancer profiles. The cumulative expression of the genes across the samples is plotted and shows a clear batch effect. After the batch QC, the values are normalized for most of the samples in all the datasets. [c] A circular dendrogram of six datasets presents a dataset-wise grouping of samples. Batch QC results in the clustering of samples across the datasets

1060 gastric tumor samples from six cohorts.

Evaluation of the batch effect criteria confirms the seamless merging

Meta-merge cohort-1060 was investigated by standard quality control analyses. Before merging, the individual cohorts were normalized by affy package to enable the comparison across datasets. The expected concern across the mRNA profiles of tumor datasets is the inherent batch effect emerging from individual datasets. To get rid of the batch effect across the datasets, a previously established method of non-parametric combat analysis *batch QC* was performed [13]. The batch effect, which occurs as a result of different attributions associated with different datasets, was attempted to get removed.

The cumulative expression of the genes across all the samples in the dataset was found to differ across the datasets. However, upon *batch QC*, the expression pattern was found uniform across all the datasets (Figure 1b). Next, the clustering patterns of gastric tumor samples from the different cohorts were visualized as a circular dendrogram upon meta-merging. Upon batch correction, the samples were found clustered based on the gene expression and not based on the datasets (Figure 1c). Third, a total of 50 randomly taken genes were investigated for their expression pattern across the

datasets and were visualized upon hierarchical clustering. The heat map showed the dataset-based clustering before batch analysis and the removal of batch effect has resulted in a convincing clustering pattern (Figure 2a). Fourth, the outlaying samples of each of the datasets were identified and removed in the batch correction process. The median was calculated for all the samples, and after batch correction, all the median values were in the same range across datasets (Figure 2b). Fifth, the expression pattern of samples in each of the datasets showed a larger overlap within datasets, and upon batch correction, the samples were found to overlap with the samples of other datasets (Figure 3a). Sixth, the median expression values of the samples were plotted in both axes, and the distribution of samples was inferred from Principal Component Analysis [PCA]. The batch correction was found to normalize the variation and facilitate the grouping of samples from multiple cohorts (Figure 3b). Seventh, the mean, variance, skewness, and kurtosis were calculated with the z-score value of genes in all the datasets. Variations across batches were reduced significantly after *batch QC* by improving data consistency (Supplementary Figure 1a). Eighth, analyzing the variability in the data across three different groups. The median value for the batches got normalized after *batch QC* which suggest the reduced overall variability in the data. Variability in the Condition group remained stable after *batch QC* as no normal samples were used in the study (Supplementary Figure 1b). Ninth, condition effect P value across genes are 1 because no normal samples were used in the study (Supplementary Figure 2a), the batch effect P values across genes for all six datasets were found to be 0.999 after *batch QC* which indicates the batch effects are not significantly influencing gene expression (Supplementary Figure 2b). All these parameters clearly demonstrate the flawless merging of 1060 gastric tumor samples from six datasets. The meta-merged cohort-1060 seems suitable for further investigation.

EMT subtype tumors represent a unique cluster among gastric tumors

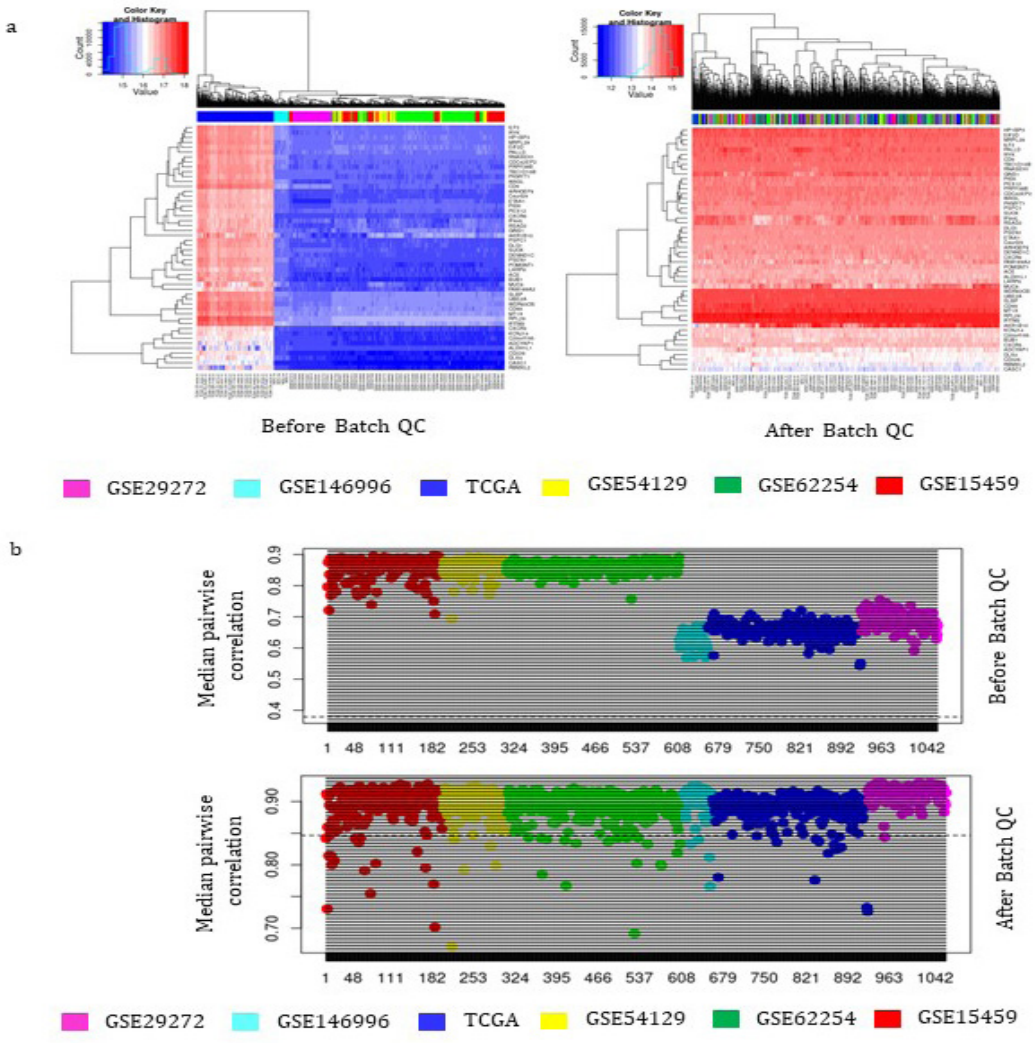


Figure 2. Evaluation of the Merging of Datasets in Meta-Merge Cohort-1060. [a] A total of 50 randomly selected genes were analyzed for their expression and clustering pattern across the datasets. The heat map shows that before batch analysis, clustering was based on datasets. After *batch QC*, the heat map shows the clear clustering of genes, irrespective of the datasets. [b] The batch effect shows the differences in the median of datasets. The median was calculated for all the samples, and the values were plotted. The median values were found to be dataset specific. Batch QC removes the outlying samples, and the median values were found to be nearly uniform across the datasets

With the merged cohort of 1,060 gastric tumor samples, gene signature-based pathway activation analysis was performed. ACRG [Asian Cancer Research Group] and TCGA [The Cancer Genome Atlas] study-based gene sets representing molecular subtypes of gastric cancer were analyzed [26,28]. Hierarchical clustering of the Z scores of the expression of the above gene sets revealed 9 clusters with differing groups of tumors with varying clinical features (Figure 4 and Supplementary Table 2). Molecular features of the subgroups of tumors were analyzed for all the clusters. Notably, from the ACRG cohort, Epithelial to Mesenchymal Transition [EMT] tumors were found clustered in cluster 9 (Figures 4a and 4b). Samples clustered along with ACRG-EMT samples with the closely associated expression pattern of the EMT gene set might reciprocate the clinical features similar to the EMT samples in other gastric tumor cohorts as well. While previously defined 40 EMT tumors of ACRG are clustered together, 165 other tumors from 1060 samples were found to cluster together and hence could be extrapolated for similar features. Notably, in cluster 9, out of 165 samples, 143 samples were with a Z score range of 0.7 – 3, which is comparable to that of the clinically validated EMT samples of the ACRG cohort. Out of 205 tumors in cluster 9, with available clinical information, most of the tumors were poorly differentiated, diffuse, and later-stage tumors, which are the known characteristic features of EMT subtype tumors [Supplementary Table 2]. A few samples in cluster 1 and cluster 2 also showed higher values for the expression of the EMT gene set. Thus, from meta-merge, it is possible to extrapolate the clinical outcome/features of a defined cohort to other cohorts. Molecular subtype-based patient stratification and treatment is evolving as a critical part of the precision medicine of cancers [29]. Meta-merging of gastric tumor cohorts seems helpful in identifying the subtype of samples. This, in turn, would facilitate the diagnosis of EMT gastric tumors and eventually treating with EMT-specific drugs.

EMT and MSI subtypes show mutual exclusivity across 90 % of gastric tumors

The Microsatellite Instability [MSI] subtype is known for favorable prognosis, early-stage tumors, and intestinal

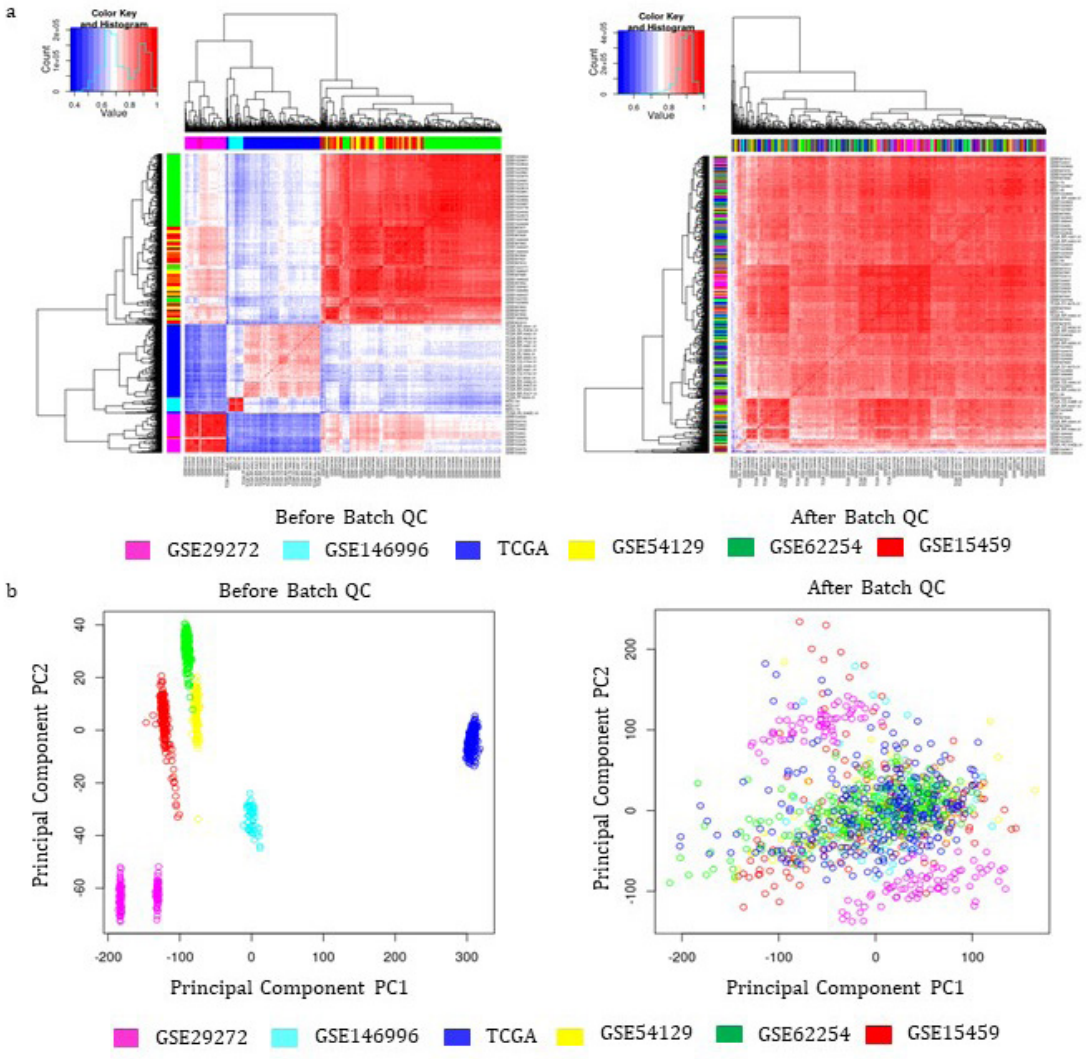
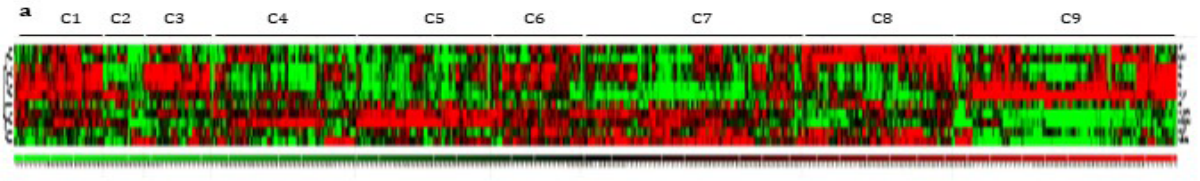


Figure 3. Parameters Confirming the Merging of Datasets in Meta-Merge Cohort-1060. [a] Expression patterns show an overlap within the datasets before *batch QC* and after *batch QC*, samples got clustered across datasets. [b] The median value of each sample was plotted on both axes. The distribution of samples in the PCA plot shows mingled distribution after batch correction

a



b

Molecular Subtypes	Cluster	C1	C2	C3	C4	C5	C6	C7	C8	C9
	Total samples	81	37	61	132	123	82	202	137	205
ACRG	EMT	0	3	0	0	0	0	1	2	40
	MSI	3	2	6	3	2	6	22	22	2
	TP53+	3	2	1	17	13	4	23	14	2
	TP53-	13	2	5	29	19	5	20	8	6
TCGA	EBV	0	0	0	2	0	0	4	17	1
	MSI	2	1	2	3	1	14	22	12	2
	CIN	14	10	9	21	11	18	22	9	14
	GS	10	1	2	5	1	2	3	1	29

c

Z score (cut of >1)										
Molecular Subtypes	Cluster	C1	C2	C3	C4	C5	C6	C7	C8	C9
	Total samples	81	37	61	132	123	82	202	137	205
ACRG	EMT	13	9	1	0	0	0	4	0	157
	MSI	2	5	8	6	4	19	53	46	8
	TP53+	7	8	5	17	20	16	25	30	27
	TP53-	12	5	13	15	13	10	37	14	30
TCGA	EBV	16	0	10	6	8	13	18	82	19
	MSI	33	0	33	4	6	21	21	17	41
	CIN	32	0	40	10	4	21	15	6	44
	GS	44	0	30	1	0	5	1	2	93

Figure 4. Distribution of Gastric Tumor Molecular Subtypes Across 1060 Samples of Meta-Merged Dataset. Based on hierarchical clustering, the samples in the meta-merged cohort 1060 were distributed into 9 clusters. Clinical subtypes of ACRG and TCGA gastric tumor cohorts were analyzed. [a] Heatmap shows the gene set-based activation pattern of ACRG and TCGA subtypes. The heatmap is provided in Supplementary Table 1 for clear visibility. [b] Table shows the number of samples with ACRG and TCGA subtypes in different clusters of meta merged cohort of 1060 gastric tumor samples. [c] Occurrence of ACRG and TCGA subtypes in different clusters of the meta-merged cohort 1060 based on the gene set-based Z score cut off > 1. The above data reveals the following: i] Occurrence of EMT in a distinct cluster, ii] exclusivity between EMT and MSI in 90 % of tumors, iii] overlap of EMT and MSI subtypes in 10 % of samples, iv] overlap and differences between EMT and GS subtype, and v] association between EBV and MSI subtype tumors.

subtype tumors. EMT subtype is characterized to show the worst prognosis, high recurrence rate, and diffuse subtype tumors [30–32]. MSI tumors are characterized by the loss of function of genes involved in mismatch repair [MMR], and EMT tumors are known for the features of alterations in cell adhesion, angiogenesis, and motility [26]. As per the classification in the ACRG study, gastric tumors were classified into EMT and MSI subtypes with two differing gene expression patterns [30, 33, 34]. Meta-merged 1060 gastric tumor samples show mutual exclusivity between EMT and MSI subtypes in 90 % of tumors. A predominant occurrence of EMT tumors was observed in cluster 9 (Figures 4 and 5a), whereas MSI subtype tumors were enriched in clusters 6, 7, and 8 (Figures 4 and 5b). A few MSI subtype samples were also observed in other clusters. However, a prominent mutual exclusivity was observed between EMT and MSI subtype tumors. Based on clinical features, EMT and MSI tumors are different from one another. With the Z score cut off > 1, out of 323 samples in both EMT and MSI subtypes, 37 samples showed overlapping features of both. This overlapping occurrence indicates that while 90 % of EMT and MSI subtype tumors are exclusive, the overlapping features occur in 10 % of gastric tumors. This study could be the first to report the overlapping occurrence of EMT and MSI subtypes in 10 % of gastric tumors. Segregation of EMT and MSI tumors with gene set-based pathway activation analysis from the merged cohorts is a good indicator for the broader stratification of gastric cancer patients.

GS subtype shows overlapping and non-overlapping features with the EMT subtype

There are similarities and differences between ACRG and TCGA classifications. The EMT subtype of ACRG has been reported to be closely related to the GS [genomically stable] subtype but with some distinct features. EMT and GS subtype fall under the diffuse subtype with the worst prognosis in terms of survival [35, 36]. In the meta-merged cohort, among ACRG and TCGA cohort samples, about 50 % of GS subtype samples occur in cluster 9,

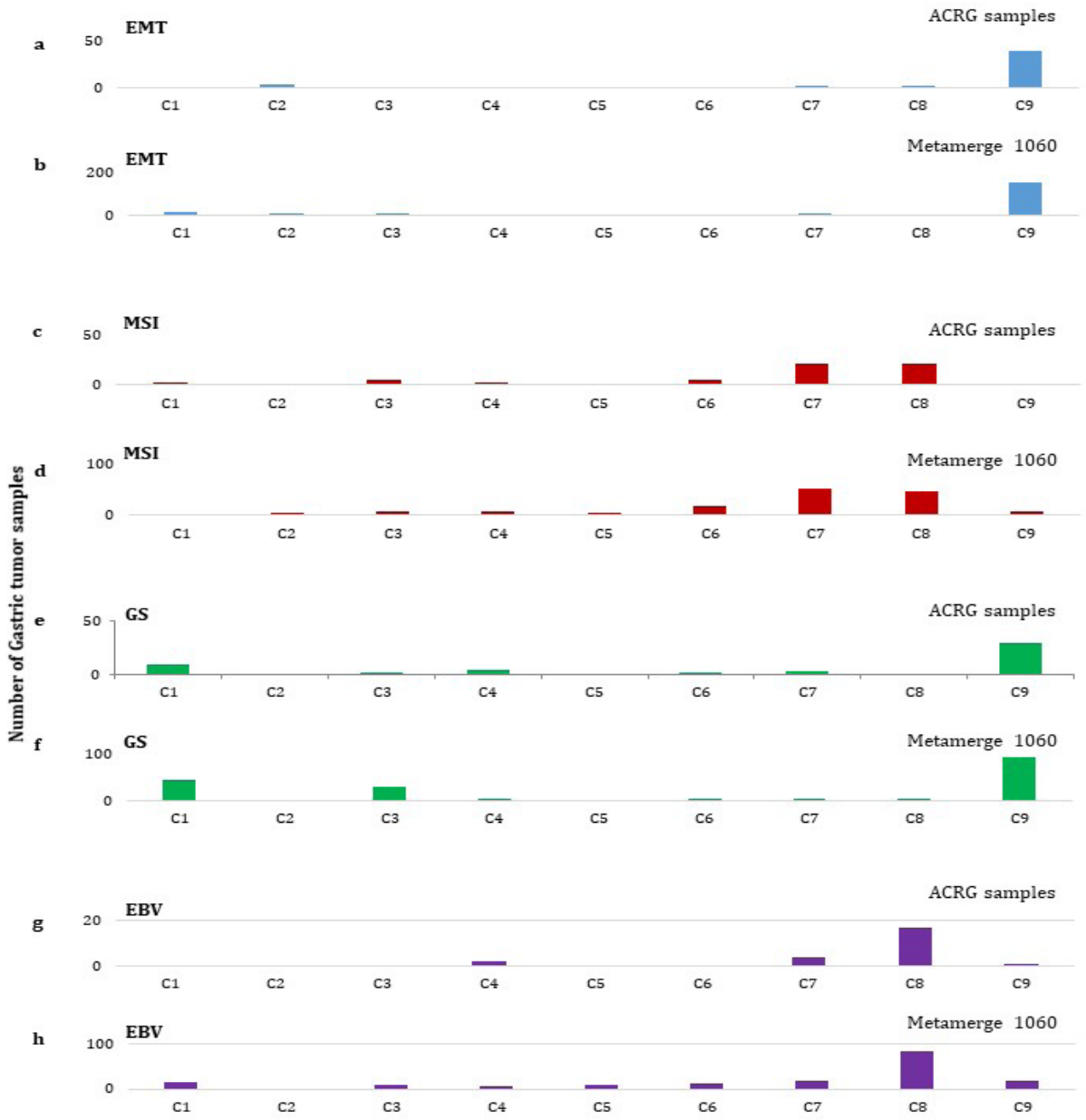


Figure 5. EMT and MSI Subtypes Show Exclusivity Across the Merged Gastric Cancer Cohort. The meta-merged 1060 gastric tumor sample cohort was analyzed for the activation pattern of the molecular subtypes defined by ACRG and TCGA studies by gene set-based Z scoring. The hierarchical clustering reveals 9 clusters of samples. [a] The occurrence of EMT subtypes as defined in the ACRG cohort by clinical and molecular methods was investigated. EMT subtype samples of the ACRG cohort are enriched in cluster 9. [b] EMT samples as defined by gene set based Z score cut off > 1 got largely enriched in cluster 9. [c-d] In the meta merged cohort - 1060, MSI subtypes are clustered in cluster 7 and cluster 9 for ACRG cohort samples [c] as well as for all samples [d]. [e-f] The majority of GS samples are grouped in cluster 8. [g-h]. Similarly, the majority of EBV samples got clustered in cluster 8 though a few are in clusters 4, 7, and 9. EMT and GS subtypes have overlapping patterns, whereas EMT and MSI show exclusivity

and 87 % of EMT samples occur in the same cluster. This reiterates the GS subtype to exhibit overlapping and non-overlapping features with the EMT subtype (Figures 4 and 5). Considering the observed differences in the clinical features, GS tumors can be classified into GS-EMT and GS-non EMT, wherein GS tumors in cluster 9 are GS-EMT tumors, and GS tumors in cluster 1 and cluster 3 are GS-non EMT. Notably, the cluster 3 does not have EMT samples. GS-EMT tumors are predominantly diffuse and poorly differentiated, whereas GS-non-EMT tumors include predominantly intestinal followed by diffuse type tumors and moderately differentiated tumors. GS-EMT and GS-non-EMT tumors being staged 2/3/4 [Supplementary Table 2]. From the current analysis, TP53, ARID1A, and KRAS mutations were observed to be comparatively higher in GS-non-EMT tumors than in GS-EMT tumors [Supplementary Table 3]. Notably, CDH1 and RHOA mutations were reported to be common in the GS subtype whereas rare in the EMT subtype [37, 38]. The current study shows a potential sub-classification scheme for EMT and GS tumors.

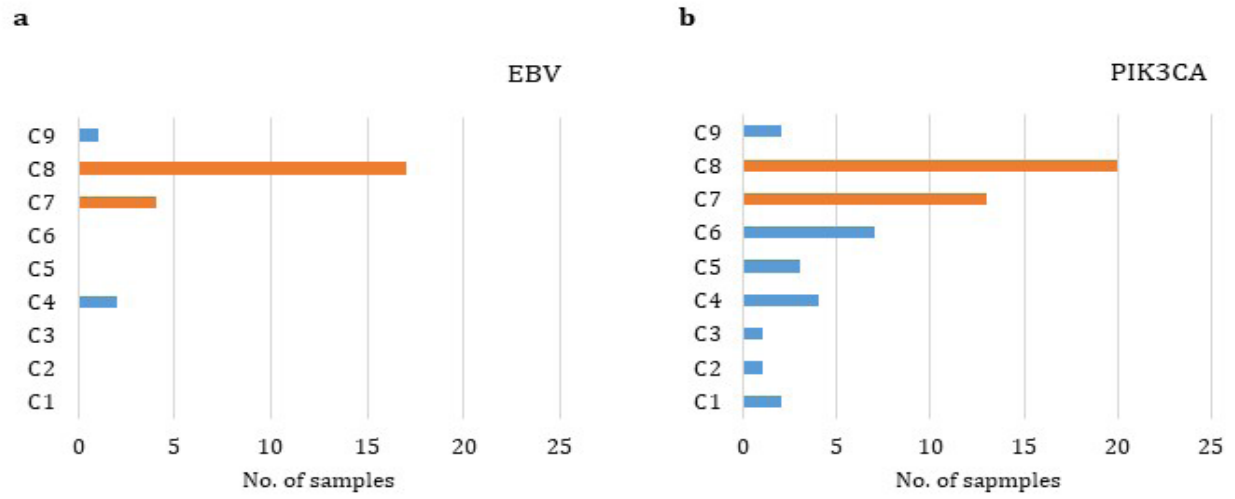


Figure 6. Association between EBV Subtype Tumors and PI3K Mutation. [a] In the meta merge 1060 cohort across TCGA cohort samples, EBV samples are to be enriched in cluster 8 and cluster 7. [b] In the TCGA cohort, the number of samples with PI3K mutations is higher in cluster 8 and cluster 7, where EBV samples were clustered. This shows the association between EBV subtype and PI3K mutation

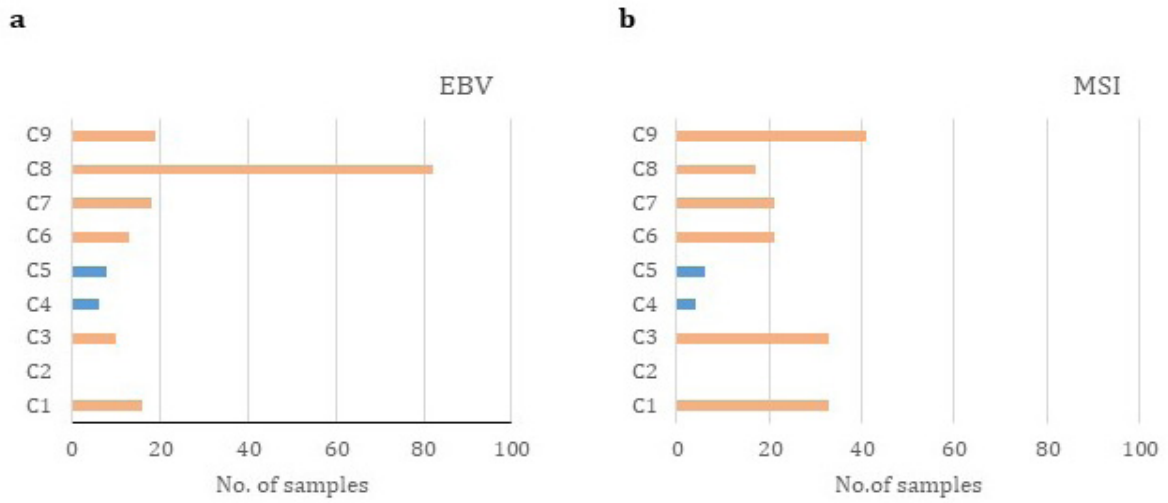


Figure 7. Association between EBV and MSI Subtype Tumors. [a] Occurrence of EBV samples [Z score cut off >1] in the meta merged cohort – 1060. Samples are largely distributed in clusters 1, 3, 6, 7, 8, and 9. [b] Distribution of MSI subtype samples [Z score cut off >1] in the meta-merged cohort - 1060. MSI subtype tumors are also enriched in clusters 1, 3. 6. 7, 8, and 9, where EBV tumors are enriched. This indicates the association between EBV and MSI subtypes of gastric tumors

Association between PI3K mutation and EBV subtype tumors

EBV [Epstein Bar Virus] plays a pivotal role in the pathogenesis of gastric cancer, and 10 % of gastric tumors are reported to be classified as EBV subtype tumors [39, 40]. A merged gastric tumor cohort with 1060 samples revealed a distinct cluster of EBV-positive tumors. Out of 24 EBV-positive tumor samples in the TCGA cohort, 17 samples fall in cluster 8 (Figure 4a-4c and 5g-5h). While the majority of EBV samples are in cluster 8, some samples were found to occur in different clusters. Further, from all the cohorts in cluster 8 with 137 samples, by referring to the Z score [cut off > 1] of the EBV gene set, 82 samples were EBV-positive. Out of 17 EBV samples of the TCGA cohort in cluster 8, 15 samples had the Z score cut off > 1, whereas across all clusters in meta-merge 1060, 172 samples were predicted to be EBV positive. This indicates that 16 % of gastric tumors are EBV positive, as inferred from the expression of the gene set, and remains to be investigated.

Analysis of PIK3CA mutation in gastric tumors from the TCGA cohort showed a correlation between PI3K mutation and EBV positivity. Meta-merged cohort showed clustering of EBV-positive tumors and PI3K mutated samples in cluster 8 followed by clusters 7, 4, and 9 (Figure 6). This shows the association between EBV and PIK3CA. In Korean and German cohorts of gastric tumors, 25 % and 32 % EBV-positive tumors showed PI3K mutations [41,

42]. All the above data shows that from meta-merging, it is possible to predict the tumor subtype and the associated molecular clinical features based on the expression of subtype-specific gene sets.

Association between EBV and MSI tumors

Epstein-Barr virus [EBV] positivity and Microsatellite Instability [MSI] are the molecular subtypes of TCGA that need to be explored for their unique characteristics and potential therapeutic implications. EBV-positive gastric tumors are poorly differentiated and associated with male patients, whereas MSI subtype gastric tumors are common in women and older patients [43–45]. EBV and MSI subtype tumors were found to co-occur in clusters – 1, 3, 6, 7, 8, and 9 (Figure 4c, 5d, 5h, and 7). Out of 294 samples with a Z-score cut-off >1 for EBV and MSI signatures, 54 samples showed an overlap between the two subtypes. This clearly indicates the overlapping molecular and possible clinical features between EBV and MSI tumors. Notably, MSI and EBV tumors were reported to be responsive to immune checkpoint blockade [43, 46, 47], and these gene sets could be used as biomarkers to identify the target patients for immune checkpoint blockade. PD-L1 expression was observed in approximately 50% and 33% of tumor cells in EBV and MSI-H tumors, respectively [48]. A 100 % and 85.7 % success rate were reported for patients with metastatic EBV and MSI treated with pembrolizumab, respectively [49]. The tumor microenvironment of EBV and MSI tumors was identified to have abundant lymphocytic infiltration [46, 47, 50]. In an independent work, we have identified the suitability of targeting EBV and MSI tumors with the proteasomal inhibitor bortezomib [communicated]. Both subtypes were reported to be associated with somatic mutations of PIK3CA and ARID1A [29, 47, 51]. Considering these reports, the identified association between EBV and MSI tumors is making sense.

The potential patient stratification and targeted therapeutics strategies emerging from the current study

From this study, i] mutual exclusivity between EMT and MSI subtypes in 90 % of gastric tumors, ii] 10 % of gastric tumors revealing the features of both EMT and MSI subtype gastric tumors, iii] classification of GS subtype tumors as GS-EMT and GS-non-EMT tumors, iv] an association between PI3K mutation and EBV subtype gastric tumors, and v] an association between EBV and MSI subtype gastric tumors were identified.

First, exclusivity between MSI and EMT subtypes in 90% of gastric tumors with 10 % of gastric tumors to exhibit the features of both EMT and MSI subtypes was identified from the current study. EMT mechanism is involved in triggering resistance of cancer cells to many chemotherapeutic agents [52, 53]. Several small molecules were identified as EMT inhibitors and to enhance chemo-sensitivity in combination with chemotherapies [54]. An array of drugs investigated for EMT and other subtypes of tumors are shown in Supplementary Table 4. For patient stratification and to enable the targeted therapeutics, the meta-merge based stratification would facilitate the diagnosis and selecting suitable therapeutics. For MSI tumors, PARP inhibitors and immune checkpoint inhibitors show improved prognosis [32]. The gene set based stratification of gastric tumors would enable identifying the overlap between EMT, MSI, and EMT/MSI tumors, which would demarcate the patients those might get benefited from targeted and combinatorial therapy.

Second, current study also reveals a potential sub-classification of GS tumors into GS-EMT and GS-non-EMT tumors. GS subtype is characterized by diffuse phenotype, advanced tumor stages, somatic alterations in genes CDH1, RHOA, and ARID1A, as well as CLDN18 and ARHGAP26 gene fusion [30, 55]. GS-EMT tumors are observed to have poorly differentiated and diffuse tumors in predominant, while GS-non-EMT tumors are mainly intestinal and moderately differentiated. This sub-classification could guide treatment choices, as these groups have differing clinical features. GS-non-EMT tumors were found to have higher rates of TP53, ARID1A, and KRAS mutations [Supplementary Table 3].

Third, the present study also reveals the occurrence of PIK3CA mutations in EBV tumors. They are also reported to have high PD-L1/PD-1 expression and strong link to PIK3CA mutation suggesting that these tumors could potentially be targeted with PI3K inhibitors. Immune checkpoint inhibitors, histone deacetylase [HDAC] inhibitors, proteasome inhibitors, indoleamine 2,3-dioxygenase inhibitors, and small-molecule EBNA1 inhibitors were reported suitable for EBV tumors [Supplementary Table 4]. In addition, PIK3CA inhibitors seem potential candidates and could be considered for combinatorial therapeutics.

Fourth, EBV and MSI subtypes were found to co-occurr in gastric tumors. A prior study has reported success rates of 100% and 85.7% for metastatic EBV and MSI tumors, respectively, upon treatment with pembrolizumab. Despite distinct clinicopathological characteristics and due to pro inflammatory immune microenvironment, both EBV and MSI tumors can be benefited by immunotherapy [50]. PD-L1 positivity was also observed in in EBV and MSI positive tumors [56]. The current study reiterates an association between EBV and MSI subtypes tumors. There are many drugs that are reportedly common for both EBV and MSI tumors, such as pembrolizumab and nivolumab [Supplementary Table 4]. Further preclinical and clinical exploration of the current classification scheme and therapeutic guidance would add significant improvement in clinical management of gastric cancers.

Discussion

The current study aimed to explore the usefulness of the meta-merging of gastric cancer transcriptome profiles and to extrapolate the molecular and clinical features across cohorts. This would also be useful to increase the statistical power of analysis. The expression profiles of different cohorts are considered non-comparable and heterogeneous mainly due to the platform differences. The successful development of meta-merging is a significant step forward [57]. The batch effect is a concern and would have an impact on pattern detection among the samples,

clustering, dimension reduction, and construction of networks between subjects [58, 59]. To overcome the batch effect, BatchQC was used [13]. ComBat tool has been reported to remove systematic technical variations which are inherent to the datasets. In the present study, six cohorts were merged, and the removal of the batch effect was evident from the comparative analysis of various parameters.

Merging of breast cancer datasets from 29K and Agilent platforms has revealed, combat, a statistical method to improve the reproducibility of the gene expression measurements [57]. The other parametric and non-parametric methods employed include adjusting data for batch effects [3, 15]. When the sources of batch effect are unknown, the SVA [Surrogate Variable Analysis] method was suggested to correct the batch effects from high throughput data [16]. Despite the availability of a few methods, including SVA, Combat, Combat, and scaling of batch effect clearance, ComBat analysis is highly used in many studies [8]. Batch QC also constructs interactive outputs for transcriptomic data to visualize and understand the technical variations.

Heterogeneity across gastric tumors in cohorts is identified by genomic and transcriptome analyses [60]. Classification of tumors into molecular subtypes by specific genetic aberrations and expression signatures are helpful in understanding biological and clinical differences [61]. Upon the learning from ACRG and TCGA studies, attempts to classify gastric tumors into molecular subtypes were made in different populations [62–65]. Efforts also were made to effectively predict the response of different molecular subtypes to systemic therapy [66]. However, these classifications and clinical features need to be investigated in multiple cohorts. In the current study, both ACRG and TCGA subtype-specific gene sets were analyzed for their expression pattern in the meta-merged cohort 1060. Extrapolation of clinical information from the ACRG cohort revealed EMT and MSI subtypes to be mutually exclusive in 90 % of tumors; however, 10% of the tumors were found to have the transcriptomic features of EMT and MSI subtype gastric tumors. MSI subtype tumors were reported to have a better prognosis, whereas EMT tumors are known to have a worse prognosis [30, 33]. Earlier analysis of differentially expressed genes in MSI and EMT subtypes did not show any overlap [34].

Upon extrapolating the clinical knowledge of TCGA cohort to other cohorts the commonality and differing aspects of GS and EMT subtypes were inferred. Notably, the current study defines GS tumors into GS-EMT tumors and GS-non-EMT tumors for the first time. Further, the association between EBV and MSI subtypes and their linkage with PIK3CA mutation were all readily visible. The current analyses show the applications of meta-merging toward extrapolating the clinical features across cohorts. However, this also depends on the gene sets. While EMT, MSI, and GS gene sets are clearly predictive of the subtypes, TP53+, TP53-, and CIN gene sets were found not to define the samples clearly.

With big data–driven targeted therapies, only 5% of the patients are estimated to get benefited [67]. The current study shows the possibility of big data-driven analysis of gene expression profiles from multiple cohorts of gastric tumors. This is a step forward in extrapolating clinical information to tumor samples. With the meta-merging of gene expression studies from different parts of the world, it is possible to analyze the clinical data for the possible development of intriguing diagnostic and therapeutic strategies for cancers.

Author Contribution Statement

Balaji T Sekar and Kumaresan Ganesan has designed the work and framed the objectives. Balaji T Sekar has performed all the computational analysis and wrote the manuscript. Kumaresan Ganesan has supervised and corrected the manuscript.

Acknowledgements

This work was supported by the Department of Biotechnology [DBT], Government of India, through the Unit of Excellence [UOE] in Cancer Genetics research grant BT/MED/30/SP11290/2015 to Dr. Kumaresan Ganesan, Madurai Kamaraj University. The grant support from MKU-RUSA grant 014/RUSA/MKU/2020-2021 is acknowledged. The authors acknowledge the Gene Expression Omnibus [GEO] from NCBI and cBioPortal for TCGA for data accessibility. The instrumentation support by the DST-FIST computational facility of the School of Biological Sciences, Madurai Kamaraj University, is acknowledged.

Conflict of interest

The authors disclose no potential conflicts of interest.

References

1. Petryszyn P, Chapelle N, Matysiak-Budnik T. Gastric cancer: Where are we heading? Dig Dis. 2020;38(4):280-5. https://doi.org/10.1159/000506509.
2. Lim SB, Tan SJ, Lim WT, Lim CT. A merged lung cancer transcriptome dataset for clinical predictive modeling. Sci Data. 2018;5:180136. https://doi.org/10.1038/sdata.2018.136.
3. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical bayes methods. Biostatistics. 2007;8(1):118-27. https://doi.org/10.1093/biostatistics/kxj037.
4. Dawany NB, Tozeren A. Asymmetric microarray data produces gene lists highly predictive of research literature on multiple cancer types. BMC Bioinformatics. 2010;11:483. https://doi.org/10.1186/1471-2105-11-483.
5. Lazar C, Meganck S, Taminau J, Steenhoff D, Coletta A, Molter C, et al. Batch effect removal methods for microarray gene expression data integration: A survey. Brief Bioinform. 2013;14(4):469-90. https://doi.org/10.1093/bib/bbs037.
6. Chang LC, Lin HM, Sibille E, Tseng GC. Meta-analysis methods for combining multiple expression profiles: Comparisons, statistical characterization and an application guideline. BMC Bioinformatics. 2013;14:368. https://doi.org/10.1186/1471-2105-14-368.
7. Zhang Y, Jenkins DF, Manimaran S, Johnson WE. Alternative empirical bayes models for adjusting for batch effects in genomic studies. BMC Bioinformatics. 2018;19(1):262.

https://doi.org/10.1186/s12859-018-2263-6.
8. Lagani V, Karozou AD, Gomez-Cabrero D, Silberberg G, Tsamardinos I. A comparative evaluation of data-merging and meta-analysis methods for reconstructing gene-gene interactions. BMC Bioinformatics. 2016;17 Suppl 5(Suppl 5):194. https://doi.org/10.1186/s12859-016-1038-1.
9. Leng D, Huan C, Xie T, Liang J, Wang J, Dai H, et al. Meta-analysis of genetic programs between idiopathic pulmonary fibrosis and sarcoidosis. PLoS One. 2013;8(8):e71059. https://doi.org/10.1371/journal.pone.0071059.
10. Grigoryev DN, Dalal J, Becker ML, Ye SQ. Combined meta-analysis of systemic effects of allogeneic stem cell transplantation and systemic sclerosis. BMC Hematol. 2014;14(1):7. https://doi.org/10.1186/2052-1839-14-7.
11. Chen PF, Wang F, Zhang ZX, Nie JY, Liu L, Feng JR, et al. A novel gene-pair signature for relapse-free survival prediction in colon cancer. Cancer Manag Res. 2018;10:4145-53. https://doi.org/10.2147/cmar.S176260.
12. Varma S. Blind estimation and correction of microarray batch effect. PLoS One. 2020;15(4):e0231446. https://doi.org/10.1371/journal.pone.0231446.
13. Manimaran S, Selby HM, Okrah K, Ruberman C, Leek JT, Quackenbush J, et al. Batchqc: Interactive software for evaluating sample and batch effects in genomic data. Bioinformatics. 2016;32(24):3836-8. https://doi.org/10.1093/bioinformatics/btw538.
14. Gagnon-Bartsch JA, Speed TP. Using control genes to correct for unwanted variation in microarray data. Biostatistics. 2012;13(3):539-52. https://doi.org/10.1093/biostatistics/kxr034.
15. Chen C, Grennan K, Badner J, Zhang D, Gershon E, Jin L, et al. Removing batch effects in analysis of expression microarray data: An evaluation of six batch adjustment methods. PLoS One. 2011;6(2):e17238. https://doi.org/10.1371/journal.pone.0017238.
16. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet. 2010;11(10):733-9. https://doi.org/10.1038/nrg2825.
17. Luo J, Schumacher M, Scherer A, Sanoudou D, Megherbi D, Davison T, et al. A comparison of batch effect removal methods for enhancement of prediction performance using maqc-ii microarray gene expression data. Pharmacogenomics J. 2010;10(4):278-91. https://doi.org/10.1038/tpj.2010.57.
18. Scherer A. Variation, Variability, Batches and Bias in Microarray Experiments: An Introduction. In: Batch Effects and Noise in Microarray Experiments [Internet]. John Wiley & Sons, Ltd; 2009 [cited 2025 Jan 2]. p. 1–4. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470685983.ch1
19. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28(6):882-3. https://doi.org/10.1093/bioinformatics/bts034.
20. Stein CK, Qu P, Epstein J, Buros A, Rosenthal A, Crowley J, et al. Removing batch effects from purified plasma cell gene expression microarrays with modified combat. BMC Bioinformatics. 2015;16:63. https://doi.org/10.1186/s12859-015-0478-3.
21. Müller C, Schillert A, Röthemeier C, Trégouët DA, Proust C, Binder H, et al. Removing batch effects from longitudinal gene expression - quantile normalization plus combat as best approach for microarray transcriptome data. PLoS One. 2016;11(6):e0156594. https://doi.org/10.1371/journal.pone.0156594.
22. Li S, Gao J, Xu Q, Zhang X, Huang M, Dai X, et al. A signature-based classification of gastric cancer that stratifies tumor immunity and predicts responses to pd-1 inhibitors. Front Immunol. 2021;12:693314. https://doi.org/10.3389/fimmu.2021.693314.
23. Gwee YX, Chia DKA, So J, Ceelen W, Yong WP, Tan P, et al. Integration of genomic biology into therapeutic strategies of gastric cancer peritoneal metastasis. J Clin Oncol. 2022;40(24):2830. https://doi.org/10.1200/jco.21.02745.
24. Joshi SS, Badgwell BD. Current treatment and recent progress in gastric cancer. CA Cancer J Clin. 2021;71(3):264-79. https://doi.org/10.3322/caac.21657.
25. Bass AJ TV, Shmulevich I, Reynolds SM, Miller M, Bernard B, et al. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014;513(7517):202-9. https://doi.org/10.1038/nature13480.
26. Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med. 2015;21(5):449-56. https://doi.org/10.1038/nm.3850.
27. Li C, Wong WH. DNA-Chip Analyzer [dChip]. In: Parmigiani G, Garrett ES, Irizarry RA, Zeger SL, editors. The Analysis of Gene Expression Data: Methods and Software [Internet]. New York, NY: Springer; 2003 [cited 2024 Sept 6]. p. 120–41. Available from: https://doi.org/10.1007/0-387-21679-0_5
28. Gulley ML. Genomic assays for epstein-barr virus-positive gastric adenocarcinoma. Exp Mol Med. 2015;47(1):e134. https://doi.org/10.1038/emm.2014.93.
29. Li B, Zhang F, Niu Q, Liu J, Yu Y, Wang P, et al. A molecular classification of gastric cancer associated with distinct clinical outcomes and validated by an xgboost-based prediction model. Mol Ther Nucleic Acids. 2023;31:224-40. https://doi.org/10.1016/j.omtn.2022.12.014.
30. Garattini SK, Basile D, Cattaneo M, Fanotto V, Ongaro E, Bonotto M, et al. Molecular classifications of gastric cancers: Novel insights and possible future applications. World J Gastrointest Oncol. 2017;9(5):194-208. https://doi.org/10.4251/wjgo.v9.i5.194.
31. Heo YJ, Park C, Yu D, Lee J, Kim KM. Reproduction of molecular subtypes of gastric adenocarcinoma by transcriptome sequencing of archival tissue. Sci Rep. 2019;9(1):9675. https://doi.org/10.1038/s41598-019-46216-6.
32. Duan Y, Xu D. Microsatellite instability and immunotherapy in gastric cancer: a narrative review. Precision Cancer Medicine. 2023 Jun 30;6.
33. Cisło M, Filip AA, Arnold Offerhaus GJ, Ciseł B, Rawicz-Pruszyński K, Skierucha M, et al. Distinct molecular subtypes of gastric cancer: From laurén to molecular pathology. Oncotarget. 2018;9(27):19427-42. https://doi.org/10.18632/oncotarget.24827.
34. Sadeghi M, Karimi MR, Karimi AH, Ghorbanpour Farshbaf N, Barzegar A, Schmitz U. Network-based and machine-learning approaches identify diagnostic and prognostic models for emt-type gastric tumors. Genes (Basel). 2023;14(3). https://doi.org/10.3390/genes14030750.
35. Nshizirungu JP, Bennis S, Mellouki I, Sekal M, Benajah DA, Lahmidani N, et al. Reproduction of the cancer genome atlas (tcga) and asian cancer research group (acrg) gastric cancer molecular classifications and their association with clinicopathological characteristics and overall survival in moroccan patients. Dis Markers. 2021;2021:9980410. https://doi.org/10.1155/2021/9980410.
36. Furukawa K, Hatakeyama K, Terashima M, Nagashima T, Urakami K, Ohshima K, et al. Molecular classification of gastric cancer predicts survival in patients undergoing radical gastrectomy based on project hope. Gastric Cancer.

2022;25(1):138-48. https://doi.org/10.1007/s10120-021-01242-0.
37. Chia NY, Tan P. Molecular classification of gastric cancer. Ann Oncol. 2016;27(5):763-9. https://doi.org/10.1093/annonc/mdw040.
38. Ooki A, Yamaguchi K. The dawn of precision medicine in diffuse-type gastric cancer. Ther Adv Med Oncol. 2022;14:17588359221083049. https://doi.org/10.1177/17588359221083049.
39. Jeong Y, Cho CE, Kim JE, Lee J, Kim N, Jung WY, et al. Deep learning model to predict epstein-barr virus associated gastric cancer in histology. Sci Rep. 2022;12(1):18466. https://doi.org/10.1038/s41598-022-22731-x.
40. Siciliano MC, Tornambè S, Cevenini G, Sorrentino E, Granai M, Giovannoni G, et al. Ebv persistence in gastric cancer cases conventionally classified as eber-ish negative. Infect Agent Cancer. 2022;17(1):57. https://doi.org/10.1186/s13027-022-00469-5.
41. Böger C, Krüger S, Behrens HM, Bock S, Haag J, Kalthoff H, et al. Epstein-barr virus-associated gastric cancer reveals intratumoral heterogeneity of pik3ca mutations. Ann Oncol. 2017;28(5):1005-14. https://doi.org/10.1093/annonc/mdx047.
42. Seo AN, Kang BW, Bae HI, Kwon OK, Park KB, Lee SS, et al. Exon 9 mutation of pik3ca associated with poor survival in patients with epstein-barr virus-associated gastric cancer. Anticancer Res. 2019;39(4):2145-54. https://doi.org/10.21873/anticanres.13328.
43. Martinez-Ciarpaglini C, Fleitas-Kanonnikoff T, Gambardella V, Llorca M, Mongort C, Mengual R, et al. Assessing molecular subtypes of gastric cancer: Microsatellite unstable and epstein-barr virus subtypes. Methods for detection and clinical and pathological implications. ESMO Open. 2019;4(3):e000470. https://doi.org/10.1136/esmoopen-2018-000470.
44. Seo JW, Park KB, Chin HM, Jun KH. Does epstein-barr virus-positive gastric cancer establish a significant relationship with the multiple genes related to gastric carcinogenesis? PLoS One. 2023;18(6):e0283366. https://doi.org/10.1371/journal.pone.0283366.
45. Li G, Zhou Z, Wang Z, Wang Z. Assessing epstein-barr virus in gastric cancer: Clinicopathological features and prognostic implications. Infect Agent Cancer. 2023;18(1):11. https://doi.org/10.1186/s13027-023-00489-9.
46. De Rosa S, Sahnane N, Tibiletti MG, Magnoli F, Vanoli A, Sessa F, et al. Ebv^{+} and msi gastric cancers harbor high pd-l1/pd-1 expression and high cd8^{+} intratumoral lymphocytes. Cancers (Basel). 2018;10(4). https://doi.org/10.3390/cancers10040102.
47. Hinata M, Ushiku T. Detecting immunotherapy-sensitive subtype in gastric cancer using histologic image-based deep learning. Sci Rep. 2021;11(1):22636. https://doi.org/10.1038/s41598-021-02168-4.
48. Rodriquenz MG, Roviello G, D'Angelo A, Lavacchi D, Roviello F, Polom K. Msi and ebv positive gastric cancer's subgroups and their link with novel immunotherapy. J Clin Med. 2020;9(5). https://doi.org/10.3390/jcm9051427.
49. Ignatova E, Seriak D, Fedyanin M, Tryakin A, Pokataev I, Menshikova S, et al. Epstein-barr virus-associated gastric cancer: Disease that requires special approach. Gastric Cancer. 2020;23(6):951-60. https://doi.org/10.1007/s10120-020-01095-z.
50. Bos J, Groen-van Schooten TS, Brugman CP, Jamaludin FS, van Laarhoven HWM, Derks S. The tumor immune composition of mismatch repair deficient and epstein-barr virus-positive gastric cancer: A systematic review. Cancer Treat Rev. 2024;127:102737. https://doi.org/10.1016/j.ctrv.2024.102737.
51. Salnikov MY, MacNeil KM, Mymryk JS. The viral etiology of ebv-associated gastric cancers contributes to their unique pathology, clinical outcomes, treatment responses and immune landscape. Front Immunol. 2024;15:1358511. https://doi.org/10.3389/fimmu.2024.1358511.
52. Ebrahimi N, Manavi MS, Faghihkhorasani F, Fakhr SS, Baei FJ, Khorasani FF, et al. Harnessing function of emt in cancer drug resistance: A metastasis regulator determines chemotherapy response. Cancer Metastasis Rev. 2024;43(1):457-79. https://doi.org/10.1007/s10555-023-10162-7.
53. Ashrafizadeh M, Mirzaei S, Hashemi F, Zarrabi A, Zabolian A, Saleki H, et al. New insight towards development of paclitaxel and docetaxel resistance in cancer cells: Emt as a novel molecular mechanism and therapeutic possibilities. Biomed Pharmacother. 2021;141:111824. https://doi.org/10.1016/j.biopha.2021.111824.
54. Du B, Shim JS. Targeting epithelial-mesenchymal transition (emt) to overcome drug resistance in cancer. Molecules. 2016;21(7). https://doi.org/10.3390/molecules21070965.
55. Kankeu Fonkoua L, Yee NS. Molecular characterization of gastric carcinoma: Therapeutic implications for biomarkers and targets. Biomedicines. 2018;6(1). https://doi.org/10.3390/biomedicines6010032.
56. Lee SM, Oh H. Association of pd-l1 positivity with epstein barr virus infection and microsatellite instability in gastric carcinomas with lymphoid stroma. Sci Rep. 2024;14(1):30932. https://doi.org/10.1038/s41598-024-81764-6.
57. Larsen MJ, Thomassen M, Tan Q, Sørensen KP, Kruse TA. Microarray-based rna profiling of breast cancer: Batch effect removal improves cross-platform consistency. Biomed Res Int. 2014;2014:651751. https://doi.org/10.1155/2014/651751.
58. Price EM, Robinson WP. Adjusting for batch effects in DNA methylation microarray data, a lesson learned. Front Genet. 2018;9:83. https://doi.org/10.3389/fgene.2018.00083.
59. Fei T, Zhang T, Shi W, Yu T. Mitigating the adverse impact of batch effects in sample pattern detection. Bioinformatics. 2018;34(15):2634-41. https://doi.org/10.1093/bioinformatics/bty117.
60. Tihagam RD, Bhatnagar S. A multi-platform normalization method for meta-analysis of gene expression data. Methods. 2023;217:43-8. https://doi.org/10.1016/j.ymeth.2023.06.012.
61. Bijlsma MF, Sadanandam A, Tan P, Vermeulen L. Molecular subtypes in cancers of the gastrointestinal tract. Nat Rev Gastroenterol Hepatol. 2017;14(6):333-42. https://doi.org/10.1038/nrgastro.2017.33.
62. Ichikawa H, Nagahashi M, Shimada Y, Hanyu T, Ishikawa T, Kameyama H, et al. Actionable gene-based classification toward precision medicine in gastric cancer. Genome Med. 2017;9(1):93. https://doi.org/10.1186/s13073-017-0484-3
63. Yoon JY, Sy K, Brezden-Masley C, Streutker CJ. Histo- and immunohistochemistry-based estimation of the tcga and acrg molecular subtypes for gastric carcinoma and their prognostic significance: A single-institution study. PLoS One. 2019;14(12):e0224812. https://doi.org/10.1371/journal.pone.0224812.
64. Wang Q, Xie Q, Liu Y, Guo H, Ren Y, Li J, et al. Clinical characteristics and prognostic significance of tcga and acrg classification in gastric cancer among the chinese population. Mol Med Rep. 2020;22(2):828-40. https://doi.org/10.3892/mmr.2020.11183.
65. Ramos M, Pereira MA, de Mello ES, Cirqueira CDS, Zilberstein B, Alves VAF, et al. Gastric cancer molecular

classification based on immunohistochemistry and in situ hybridization: Analysis in western patients after curative-intent surgery. World J Clin Oncol. 2021;12(8):688-701. https://doi.org/10.5306/wjco.v12.i8.688.
66. Shi D, Yang Z, Cai Y, Li H, Lin L, Wu D, et al. Research advances in the molecular classification of gastric cancer. Cell Oncol (Dordr). 2024;47(5):1523-36. https://doi.org/10.1007/s13402-024-00951-9.
67. Barker AD, Lee JSH. Translating "big data" in oncology for clinical benefit: Progress or paralysis. Cancer Res. 2022;82(11):2072-5. https://doi.org/10.1158/0008-5472.Can-22-0100.