

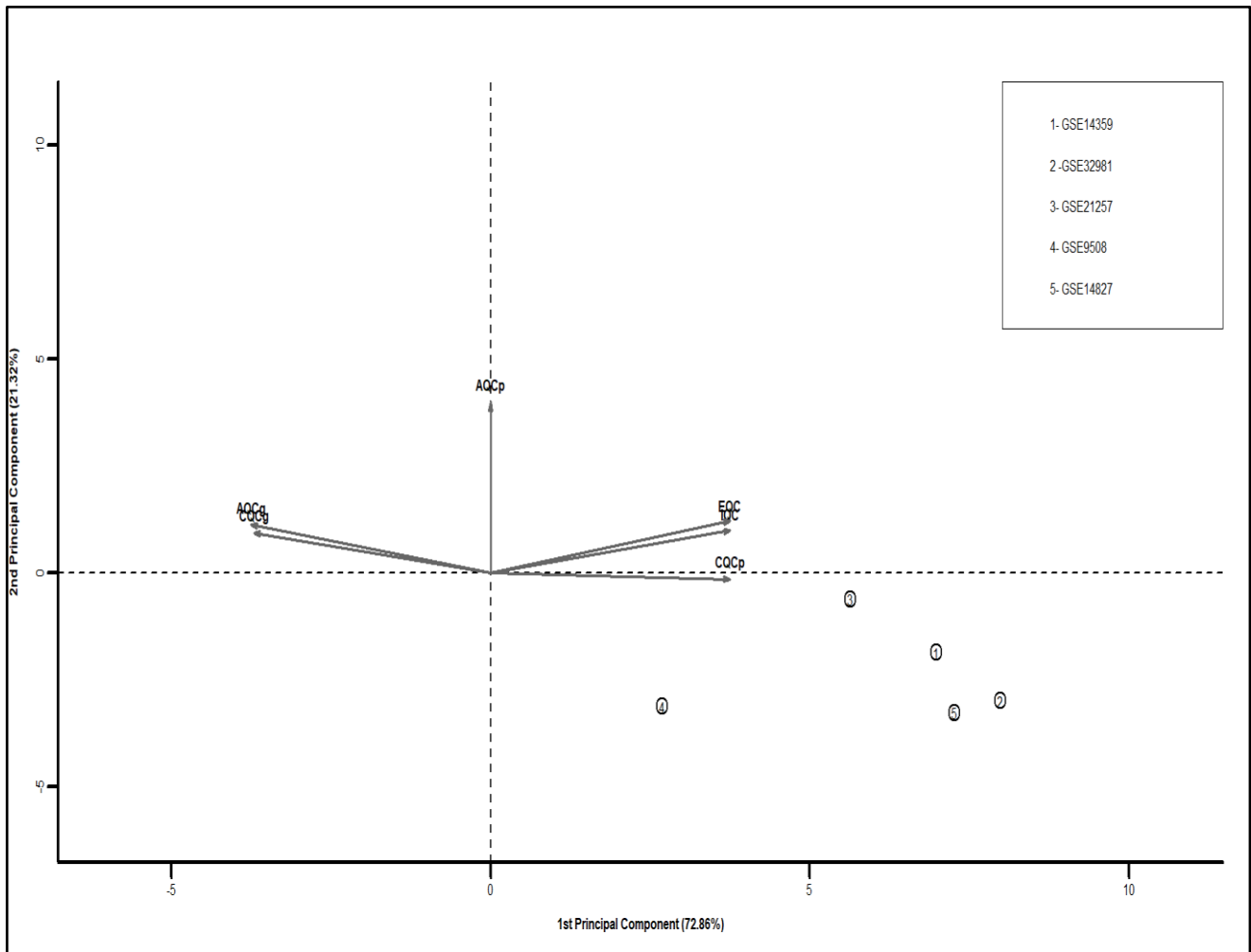
# RESULTS

## A] Quality Control Test of All Gene Expression Datasets

Study	IQC	EQC	CQCg	CQCp	AQCg	AQCp	SMR
<b>GSE32981</b>	5.8	3.26	0.01	3.5	0.08	0.14	2.58
<b>GSE21257</b>	4.14	3.1	0.42	1.63	0.45	5.92	2.67
<b>GSE14359</b>	4.5	3.26	0.38	3.28	0.14	3.03	2.42
<b>GSE14827</b>	3.79	3.35	0	2.42	0.01	0.02	3.83
<b>GSE9508</b>	0.61	0.85	0.78	0.55	0.62	0.02	3.50

**Table 1:** Quality Control of Gene Expression Datasets. **IQC**, internal quality control; **EQC**, external quality control; **CQCg**, consistency quality control of genes; **CQCp**, consistency quality control of pathways; **AQCg**, accuracy quality control of genes; **AQCp**, accuracy quality control of pathways; **SMR**, standardized mean rank.

## B] Principal Component Analysis of Gene Expression Datasets



**Figure 1:** Principal Component Analysis of Gene Expression Datasets. **PCA**, principal component analysis; **CQCG**, consistency quality control of genes; **AQCG**, accuracy quality control of genes; **IQC**, internal quality control; **CQCP**, consistency quality control of pathways; **AQCP**, accuracy quality control of pathways; **EQC**, external quality control.

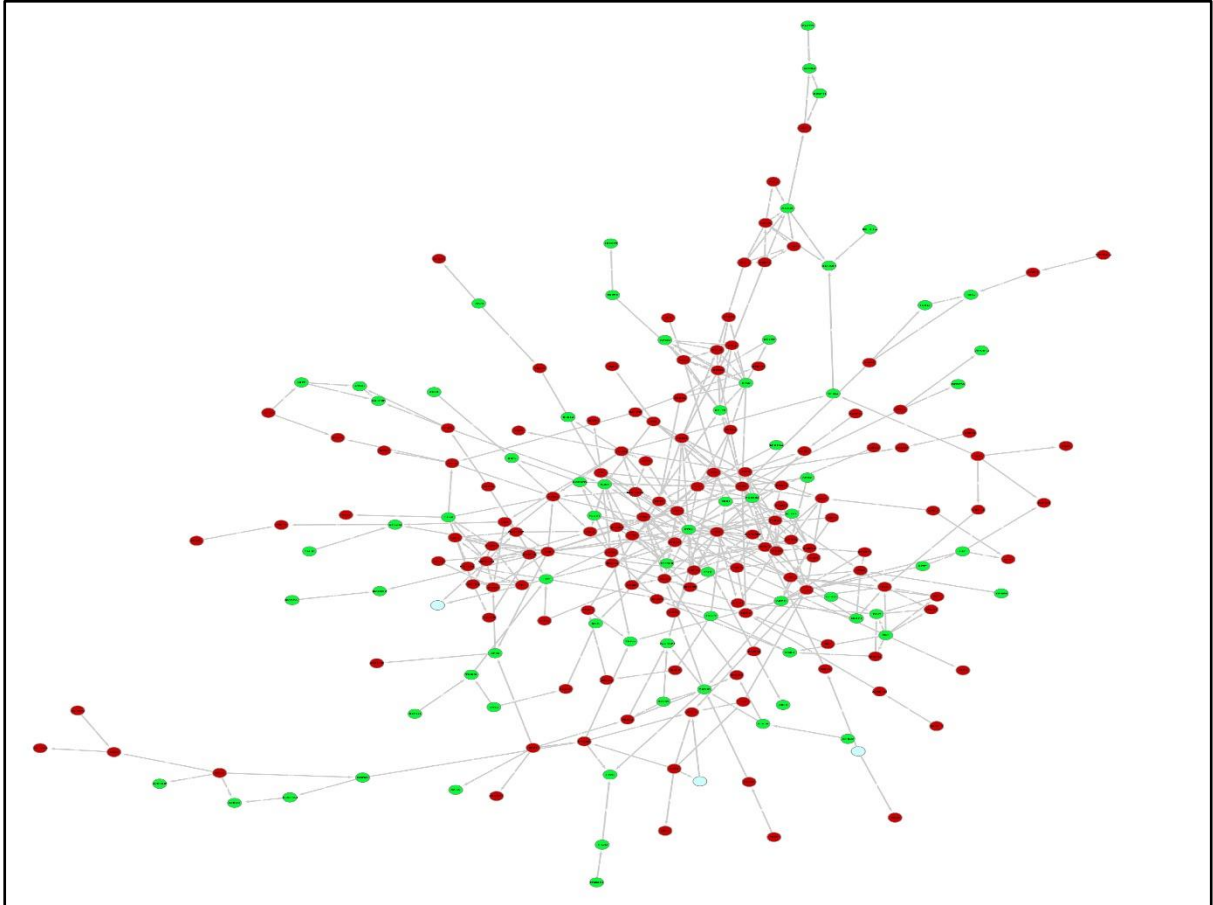
Based on these findings as well as the author's suggestion, the fifth study GSE9508 was eliminated from further analysis.

### C] Top 10 DEGs ranked by P-value

Gene	P-value	Qval	Qpval	Tau2	Expression
TREX2	7.709751e-05	1.1401465	0.7673924	0	+
IL2RA	1.667926e-04	2.3922615	0.4950768	0	-
NCOA3	2.423784e-04	0.9359168	0.8167527	0	+
PARD6A	2.534643e-04	1.0196015	0.7965091	0	-
PGRMC1	3.018393e-04	1.1253694	0.7709536	0	-
WAC	3.411439e-04	1.4663145	0.6900665	0	+
TP53I3	3.779289e-04	2.1082850	0.5502385	0	-
PP3CB	4.202570e-04	1.3439583	0.7187221	0	-
SYNJ1	4.580499e-04	2.4091393	0.4919356	0	+
ACADSB	4.600655e-04	0.6683253	0.8806276	0	-

**Table 2:** Top 10 Differentially Expressed Genes ranked by p-value and associated statistics '+' in the column 'Expression' indicates upregulation in metastatic OS compared with non-metastatic OS, while '-' indicates downregulation in metastatic OS. DEGs, differentially expressed genes; OS, osteosarcoma.

#### D] Protein-Protein Interaction Network



**Figure 2:** Protein-Protein Interaction network of the DEGs. The network comprises of 209 nodes and 435 edges. Upregulated genes in metastatic OS are shown in red, downregulated genes in metastatic OS are shown in green and non-DEGs are shown in light blue.

### E] Top 10 genes ranked by Betweenness Centrality Score

Gene	BC	EXP	Degree	p-value	Qval	Qpval	Tau2
TFRC	0.2473	+	17	0.017869	0.4440928	0.9309906	0
PTEN	0.201317	-	27	0.0047976 82	1.977504	0.5770889	0
ATM	0.152849	+	16	0.0433333 3	1.820659	0.6104496	0
LDLR	0.11639	+	10	0.0254084 2	2.049442	0.5622084	0
CYP2B6	0.088806	+	6	0.0246495 3	1.244282	0.7424044	0
PIK3C3	0.083156	+	10	0.0204303 4	2.726814	0.4356898	0
PLCB4	0.079681	+	11	0.0204887 9	1.590272	0.6615975	0
GNG10	0.078687	-	9	0.0012894 94	2.909146	0.4058462	0
SYNJ1	0.07121	+	12	0.0004580 499	2.409139	0.4919356	0
HDAC6	0.065871	+	9	0.0226062	1.393806	0.7069872	0

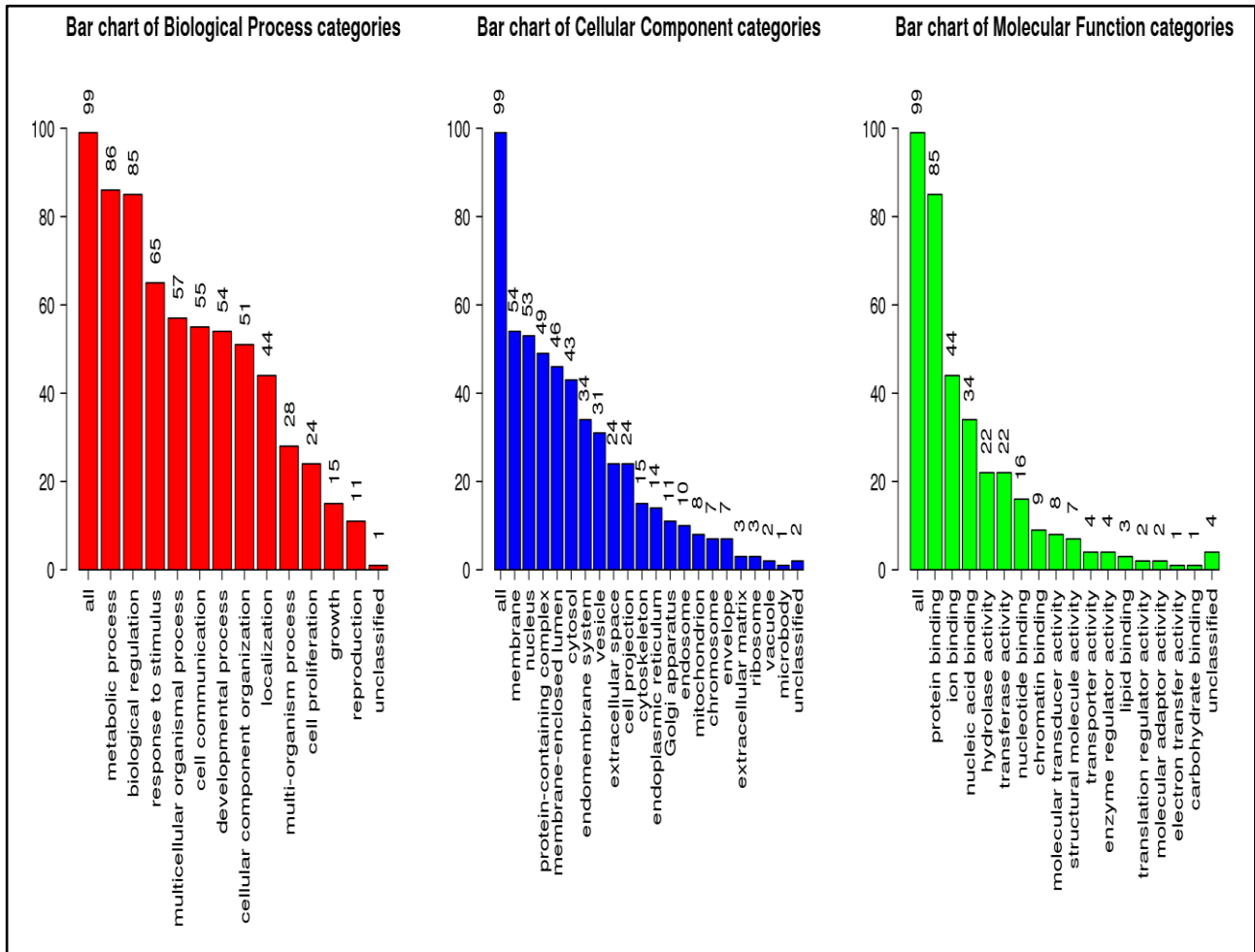
**Table 3:** '+' in the column 'EXP' indicates upregulation in metastatic osteosarcoma, while '-' indicates downregulation and 'null' indicates no differential expression.

### F] Support Vector Machine Classifier Results

Dataset	No of samples	Accuracy	Sensitivity	Specificity	PPV	NPV
Train	86	0.9302	1.0000	0.8378	0.8909	1.0000
Test	35	0.7143	0.9000	0.4667	0.6923	0.7778

**Table 4:** Prediction Results of SVM Classifier. SVM, support vector machine; PPV, positive predictive value; NPV, negative predictive value.

## G] Gene Set Enrichment Analysis



**Figure 3:** Results of Over-Representation Analysis using Gene Ontology Database and WebGestalt. Bar plots illustrating the number of genes from the analysis that overlap with the annotated genes in the GOSlim terms from biological process (red bar plot), cellular component (blue barplot) and molecular function (green bar plot) ontologies, respectively

## H] Over-Representation Analysis using WebGestalt and Gene Ontology Database

ID	Name	p-value	Hit in Query List
83061	Wnt signaling pathway	5.40E-05	WNT11,FZD8,PPP3CB,NFATC1,PLCB4,LEF1,NLK
373901	HTLV-I infection	6.16E-05	PDGFA,WNT11,CREM,FZD8,PPP3CB,NFATC1,DLG1,ATM,IL2RA
749777	Hippo signaling pathway	8.63E-05	WNT11,FZD8,PARD6A,YAP1,DLG1,PRKCI,LEF1
83105	Pathways in cancer	3.54E-04	TRAF5,PDGFA,WNT11,PTEN,FZD8,ROCK1,PLCB4,FGFR3,GNG10,LEF1
193328	mRNA surveillance pathway	3.89E-04	HBS1L,UPF3A,PCF11,UPF2,SYMPK
102279	Endocytosis	3.97E-04	LDLR,TFRC,ARF6,PARD6A,KDR,PRKCI,FGFR3,IL2RA
83052	Phosphatidylinositol signaling system	5.22E-04	CALM2,PTEN,SYNJ1,PIK3C3,PLCB4
868086	Rap1 signaling pathway	5.77E-04	PDGFA,CALM2,PARD6A,KDR,PRKCI,PLCB4,FGFR3
83092	Melanogenesis	6.28E-04	CALM2,WNT11,FZD8,PLCB4,LEF1
868085	Ras signaling pathway	9.13E-04	PDGFA,CALM2,REL,ARF6,KDR,FGFR3,GNG10
1222951	Renin secretion	1.00E-03	CALM2,ADCYAP1,PPP3CB,PLCB4
83071	Tight junction	1.07E-03	MPDZ,PARD6A,DLG1,ROCK1,PRKCI,SYMPK
83065	Axon guidance	1.25E-03	SEMA6B,TRPC3,PPP3CB,EFNB2,PARD6A,ROCK1
82987	Inositol phosphate metabolism	1.39E-03	PTEN,SYNJ1,PIK3C3,PLCB4
1404798	EGFR tyrosine kinase inhibitor resistance	2.07E-03	PDGFA,PTEN,KDR,FGFR3
1474302	Fluid shear stress and atherosclerosis	2.86E-03	PDGFA,CALM2,GSTM3,TRPV4,KDR
1435207	Breast cancer	3.03E-03	NCOA3,WNT11,PTEN,FZD8,LEF1
952859	Oxytocin signaling pathway	3.93E-03	CALM2,PPP3CB,NFATC1,ROCK1,PLCB4
852705	MicroRNAs in cancer	4.33E-03	GLS2,PDGFA,PTEN,ROCK1,FGFR3,ATM,CDC25B
983748	cGMP-PKG signaling pathway	5.14E-03	CALM2,PPP3CB,NFATC1,ROCK1,PLCB4

**Table 5:** KEGG pathways significantly over-represented in the 64 featured genes. KEGG, Kyoto Encyclopedia of Genes and Genomes.