

TCGA computational histopathology pipeline reveals subtypes and their molecular signature

Hang Chang, Ju Han, Cemal Bilgin, Gerald Fontenay, Alexander Borowski, Joe Gray, Paul Spellman, and

Bahram Parvin

Lawrence Berkeley Laboratory

Computational histopathology pipeline captures molecular basis for each morphometric subtype

BERKELEY LAB



Use case and target for analysis in the second seco

BERKELEY I

- Glioblastoma multiforme (GBM)
 - Curated by removing tissue sections with artifacts (e.g., fold in tissue, pen mark, scanning anamoly)
 - Sample size
 - 380 tissue sections selected out of 447
 - 146 patients selected out of 152
- Challenges?
 - Technical and biological variations, very large datasets
- Approach
 - Development of robust and efficient image analysis algorithms
 - Computing morphometric features and meta-features
 - Subtyping based on selected features or reduced dimensionality (e.g., PCA, MDS)
 - Molecular association with morphometric subtypes

New algorithm enhances nuclear segmentation in the presence of technical variations



Seed detection provides shape signature and local statistics



Cell-by-cell segmentation result





Cell-by-cell segmentation result





Representation

.....

AE



Normalization across all tissue sections

What are subtypes based on cellularity and nuclear size at the patient level



What is the distribution of each subtype and how well each subtype predicts survival as a function of treatment?



What are the molecular basis of each subtype?



- Gene selection
 - Univariate or multivariate methods
 - Pathway or subnetwork enrichment analysis

Subtype1 Focal Adhesion Regulation Actin Cytoskeleton Regulation CAV1,MET,ERBB4,KIT,PDGFRA,SGCE,RASA4,PDLIM3 0.000208 Subtype1 Gap Junction Regulation Adherens Junction Regulation MET,ERBB4,KIT,PDGFRA,SGCE,RASA4,PDLIM3 0.008248 Adherens Junction Regulation DAAM2,MET,ERBB4,KIT,NPY2R,PDGFRA,RASA4 0.011068 KIT -> STAT signaling KIT 0.017364 HGFR -> STAT signaling MET 0.0203089 PDGFR -> STAT signaling MET 0.0203089 PDGFR -> STAT signaling MET 0.054015 Name Overlapping Entities 0.0030127 CCR1 -> STAT signaling CCL4,CCL3 0.004022 Gap Junction Regulation Regulation GNA01,CCL4,HRH1,KIT,CL3,CALCRL,ADCY2,FGF12,RASA4 0.0083737 Subtype3 Name Overlapping Entities p-value RAT -> STAT signaling KIT 0.018322 0.0033533 Subtype4 Name Overlapping Entities p-value Name Overlapping Entities p-value RASGRP1,GNG4 0.026307 0.026307		Name	Overlapping Entities		p-value	
Subtype1 Actin Cytoskeleton Regulation MET,ERB84,KIT,PDGFRA,SGCE,RASA4,PDLIM3 0.000555 Gap Junction Regulation Adherens Junction Regulation DAAM2,MET,ERB84,KIT,NPY2R,PDGFRA,RASA4 0.008248 KIT -> STAT signaling KIT 0.011068 KIT -> STAT signaling KIT 0.017364 PDGFR -> STAT signaling MET 0.023089 PDGFR -> STAT signaling MET 0.025939 HGFR -> STAT signaling MET 0.054015 Name Overlapping Entities 9-value CCR1 -> STAT signaling CCL4,CCL3 0.0003127 Gap Junction Regulation GNA01,CCL4,HRH1,KIT,CCL3,CALCRL,ADCY2,FGF12,RASA4 0.008737 Name Overlapping Entities 9-value CCR1 -> STAT signaling KIT 0.0133533 Subtype3 Name Overlapping Entities 9-value Name Overlapping Entities 9-value Subtype4 Name Overlapping Entities 9-value Name Overlapping Entities 9-value Subtype4 Name Overlapping Entities 9-value Subtype4 Name Overlapping Entities	Subtype1	Focal Adhesion Regulation	CAV1,MET,ERBB4,KIT,PDGI	FRA,RASA4	0.000208	
Subtype1Name CCR1 -> STAT signaling Gap Junction Regulation MET_ERBB4,KIT,PDGFRA,CDH60.000555 0.0017364 0.011068KIT -> STAT signaling HGFR -> STAT signalingMET0.017364 0.023089 PDGFRAMameOverlapping Entities Gap Junction Regulation MET0.025939 0.054015Subtype3Name LCR1 -> STAT signaling METOverlapping Entities GAP Junction Regulation Overlapping EntitiesName LL11R -> STAT signaling Gap Junction Regulation KIT -> STAT signalingCCL4,CCL3 GNA01,CCL4,HRH1,KIT,CCL3,CALCRL,ADCY2,FGF12,RASA4 O.033533Subtype4Name LL11R -> STAT signaling SignalingOverlapping Entities GNA01,CCL4,HRH1,KIT,CCL3,CALCRL,ADCY2,FGF12,RASA4 O.033533		Actin Cytoskeleton			0 000555	
Subtype1 Gap Junction Regulation Adherens Junction Regulation MET,ERBB4,KIT,NPY2R,PDGFRA,RASA4 0.008248 Adherens Junction Regulation DAAM2,MET,ERBB4,KIT,PDGFRA,CDH6 0.011068 KIT -> STAT signaling KIT 0.017364 HGFR -> STAT signaling MET 0.023089 PDGFR -> STAT signaling PDGFRA 0.025939 HGFR -> FOXO3A signaling MET 0.054015 Name Overlapping Entities p-value CCR1 -> STAT signaling CCL4,CCL3 0.004022 Gap Junction Regulation GNA01,CCL4,HRH1,KIT,CCL3,CALCRL,ADCY2,FGF12,RASA4 0.008737 Subtype3 Name Overlapping Entities p-value KIT -> STAT signaling KIT 0.033533 KIT -> STAT signaling KIT 0.033533 Subtype4 Name Overlapping Entities p-value Name Overlapping Entities p-value IL11R -> STAT3 signaling IL11RA 0.018322 ThromboxaneR -> CREB RASGRP1,GNG4 0.026307		Regulation	MET,ERBB4,KIT,PDGFRA,SC	JCE, RASA4, PDLIME	3 0.000555	
Adherens Junction RegulationDAAM2,MET,ERBB4,KIT,PDGFRA,CDH60.011068KIT -> STAT signalingKIT0.017364HGFR -> STAT signalingMET0.023089PDGFR -> STAT signalingPDGFRA0.025939HGFR -> FOX03A signalingMET0.054015NameOverlapping Entitiesp-valueCCR1 -> STAT signalingCCL4,CCL30.0054015NameOverlapping Entitiesp-valueCCR5 -> TP53 signalingCCL4,CCL30.003127Gap Junction Regulation KIT -> STAT signalingGNA01,CCL4,HRH1,KIT,CCL3,CALCRL,ADCY2,FGF12,RASA40.008737KIT -> STAT signalingKIT0.018322ThromboxaneR -> CREB signalingRASGRP1,GNG40.026307		Gap Junction Regulation	MET,ERBB4,KIT,NPY2R,PDC	GFRA,RASA4	0.008248	
Regulation DAAM2,MET,ERBB4,KIT,PDGFRA,CDH6 0.011068 KIT -> STAT signaling KIT 0.017364 HGFR -> STAT signaling MET 0.023089 PDGFR -> STAT signaling PDGFRA 0.025939 HGFR -> FOXO3A signaling MET 0.054015 Name Overlapping Entities p-value CCR1 -> STAT signaling CCL4,CCL3 0.004022 Gap Junction Regulation GNAO1,CCL4,HRH1,KIT,CCL3,CALCRL,ADCY2,FGF12,RASA4 0.008737 KIT -> STAT signaling CCl4,CCL3 0.018322 Name Overlapping Entities p-value KIT -> STAT signaling IL11RA 0.018322 Subtype4 Name Overlapping Entities p-value RASGRP1,GNG4 0.026307 0.026307		Adherens Junction				
KIT -> STAT signaling KIT 0.017364 HGFR -> STAT signaling MET 0.023089 PDGFR -> STAT signaling PDGFRA 0.025939 HGFR -> FOXO3A signaling MET 0.054015 Name Overlapping Entities p-value CCR1 -> STAT signaling CCL4,CCL3 0.003127 CCR5 -> TP53 signaling CCL4,CCL3 0.004022 Gap Junction Regulation GNA01,CCL4,HRH1,KIT,CL3,CALCRL,ADCY2,FGF12,RASA4 0.008737 KIT -> STAT signaling CVerlapping Entities p-value Name Overlapping Entities p-value KIT -> STAT signaling CL1,CCL3, 0.033533 Subtype4 Name Overlapping Entities p-value Name Overlapping Entities p-value KIT -> STAT signaling IL1RA 0.018322 ThromboxaneR -> CREB RASGRP1,GNG4 0.026307		Regulation	DAAM2,MET,ERBB4,KIT,PD	GFRA,CDH6	0.011068	
HGFR -> STAT signaling MET 0.023089 PDGFR -> STAT signaling PDGFRA 0.025939 HGFR -> FOXO3A signaling MET 0.054015 Name Overlapping Entities p-value CCR1 -> STAT signaling CCL4,CCL3 0.003127 CCR5 -> TP53 signaling CCL4,CCL3 0.004022 Gap Junction Regulation GNA01,CCL4,HRH1,KIT,CCL3,CALCRL,ADCY2,FGF12,RASA4 0.008737 KIT V 0.033533 Subtype4 Name Overlapping Entities p-value Name Overlapping Entities p-value IL11R -> STAT3 signaling IL11RA 0.018322 ThromboxaneR -> CREB RASGRP1,GNG4 0.026307		KIT -> STAT signaling	KIT		0.017364	
PDGFR -> STAT signalingPDGFRA0.025939HGFR -> FOXO3A signalingMET0.054015NameOverlapping Entitiesp-valueCCR1 -> STAT signalingCCL4,CCL30.003127CCR5 -> TP53 signalingCCL4,CCL30.004022Gap Junction RegulationGNA01,CCL4,HRH1,KIT,CL3,CALCRL,ADCY2,FGF12,RASA40.008737KIT -> STAT signalingKIT0.018322NameOverlapping Entitiesp-valueIL11R -> STAT3 signalingIL11RA0.018322ThromboxaneR -> CREBRASGRP1,GNG40.026307		HGFR -> STAT signaling	MET		0.023089	
HGFR -> FOXO3A signalingMET0.054015NameOverlapping Entitiesp-valueCCR1 -> STAT signalingCCL4,CCL30.003127CCR5 -> TP53 signalingCCL4,CCL30.004022Gap Junction RegulationGNAO1,CCL4,HRH1,KIT,CCL3,CALCRL,ADCY2,FGF12,RASA40.008737KIT -> STAT signalingKIT0.018322NameOverlapping Entitiesp-valueIL11R -> STAT3 signalingIL11RA0.018322ThromboxaneR -> CREBRASGRP1,GNG40.026307		PDGFR -> STAT signaling	PDGFRA		0.025939	
NameOverlapping Entitiesp-valueCCR1 -> STAT signalingCCL4,CCL30.003127CCR5 -> TP53 signalingCCL4,CCL30.004022Gap Junction RegulationGNAO1,CCL4,HRH1,KIT,CCL3,CALCRL,ADCY2,FGF12,RASA40.008737KIT -> STAT signalingKIT0.018322NameOverlapping Entitiesp-valueIL11R -> STAT3 signalingIL11RA0.018322ThromboxaneR -> CREBRASGRP1,GNG40.026307		HGFR -> FOXO3A signaling	MET		0.054015	
Subtype3 CCR1 -> STAT signaling CCL4,CCL3 0.003127 CCR5 -> TP53 signaling CCL4,CCL3 0.004022 Gap Junction Regulation GNA01,CCL4,HRH1,KIT,CCL3,CALCRL,ADCY2,FGF12,RASA4 0.008737 KIT -> STAT signaling KIT 0.003127 Name Overlapping Entities p-value IL11R -> STAT3 signaling IL11RA 0.018322 ThromboxaneR -> CREB RASGRP1,GNG4 0.026307	Subtype3	Name	Overlapping Entities			p-value
Subtype3CCR5 -> TP53 signaling Gap Junction Regulation KIT -> STAT signalingCCL4,CCL3 GNA01,CCL4,HRH1,KIT,CCL3,CALCRL,ADCY2,FGF12,RASA4 KIT0.004022 0.008737 0.033533NameOverlapping Entities IL11R -> STAT3 signaling ThromboxaneR -> CREB signalingp-value RASGRP1,GNG40.018322 0.026307		CCR1 -> STAT signaling	CCL4,CCL3			0.003127
Gap Junction Regulation KIT -> STAT signaling GNAO1,CCL4,HRH1,KIT,CCL3,CALCRL,ADCY2,FGF12,RASA4 0.008737 Name Overlapping Entities p-value IL11R -> STAT3 signaling IL11RA 0.018322 ThromboxaneR -> CREB RASGRP1,GNG4 0.026307		CCR5 -> TP53 signaling	CCL4,CCL3			0.004022
KIT -> STAT signaling KIT 0.033533 Name Overlapping Entities p-value IL11R -> STAT3 signaling IL11RA 0.018322 ThromboxaneR -> CREB RASGRP1,GNG4 0.026307		Gap Junction Regulation	GNAO1,CCL4,HRH1,KIT,	CCL3,CALCRL,ADC	Y2,FGF12,RASA4	0.008737
NameOverlapping Entitiesp-valueIL11R -> STAT3 signalingIL11RA0.018322ThromboxaneR -> CREB signalingRASGRP1,GNG40.026307		KIT -> STAT signaling	КІТ			0.033533
NameOverlapping Entitiesp-valueIL11R -> STAT3 signalingIL11RA0.018322ThromboxaneR -> CREB signalingRASGRP1,GNG40.026307						
Subtype4 IL11R -> STAT3 signaling IL11RA 0.018322 ThromboxaneR -> CREB signaling RASGRP1,GNG4 0.026307		Name	Overlapping Entities	p-value		
Subtype4 ThromboxaneR -> CREB signaling RASGRP1,GNG4 0.026307	Subtype4	IL11R -> STAT3 signaling	IL11RA	0.018322		
signaling RASGRP1,GNG4 0.026307		ThromboxaneR -> CREB				
		signaling	RASGRP1,GNG4	0.026307		
EphrinR -> actin signaling EFNB3,SGCE,EPB41L2 0.02702		EphrinR -> actin signaling	EFNB3,SGCE,EPB41L2	0.02702		

Can tumor composition be characterized?



 Since tumor is heterogeneous, can we query for subtypes at the block levels and learn about tumor composition?



What are the tumor histology subtypes?



Does heterogeneity play a role in survival as a result of a more intense therapy ?





Another view: Are cellularity and nuclear size correlated? And outcome?



High cellularity and low nuclear size are better predictive of a more aggressive therapy

rrrr

BERKELEY

AE

Conclusion



- There are many ways to slice through the data and metadata
 - Cellularity, nuclear size
 - Heterogeneity
- Different indices lead to alternative subtypings
 - Alternative biological interpretation is possible
- Genomic association has the potential to reveal new insight
- Web site: tcga.lbl.gov
 - "Google map" like viewing of tissue sections with segmentation results overlaid