

M2GEN

Release Details for ORIEN Members

April 17, 2018

April 17th: Key Updates



- ORIEN Avatar patient Total: 3434
 - 89 new patients added
- Additional molecular data added on 6 existing patients
 - 6 patients with WES G added
- Both v2.03 schema (flat file) and normalized clinical data files are available for download
 - Please note, the flat file schema v2.03 will be phased out in 60 days (May 25th, 2018)

ORIEN Avatar: April 17th Release Summary

Patient Summary By Disease

Disease	Avatar Patient Count
BRE - Breast Cancer	482
H&N - Head and Neck Cancer	434
HEM - Multiple Myeloma	393
SAR - Sarcoma	272
GI - Colorectal Cancer	268
END - Thyroid Cancer	198
GU - Kidney Cancer	197
THO - Lung Cancer	194
GI - Pancreatic Cancer	168
GYN - Ovarian Cancer	166
GU - Prostate Cancer	160
GU - Bladder Cancer	84
GI - Other GI Cancer	71
HEM - Other HEME Disease	67
GYN - Endometrial Cancer	42
GYN - Other GYN Cancer	42
GI - Gastric Cancer	37
GI - Liver Cancer	37
CUT - Melanoma	25
GI - Esophageal Cancer	22
HEM - Leukemia	20
NEU - Brain Cancer	14
GU - Other GU Cancer	12
OTH - Other Cancer, NOS	11
HEM - Lymphoma	8
THO - Other Thoracic Cancer	3
CUT - Other Cutaneous Cancer	2
OTH - Other Non-Tumor	2
GU - Bladder Cancer and GI - Other GI Cancer	1
HEM - Leukemia and GI - Esophageal Cancer	1
GU - Prostate Cancer and GI - Pancreatic Cancer	1
Total Avatar Patients	3434

Release Deliverables

- **DNANexus (Molecular Data Host)**
 - For the additional RNA-Seq data added, files include (per sample):
 - FASTQ
 - BAM
 - BAI
 - Cufflinks
 - For the additional WES data added, files include (per sample):
 - FASTQ – tumor and normal
 - BAM – tumor and normal
 - BAI
 - VCF – tumor and normal
 - Paired Mutect2 Haplotyper.VCF
 - Paired Mutect1 stats.txt

- **sFTP Download Folder (Member Specific Folder)**
 - The following files are available for download:
 - Linkage File
 - RNA-Seq QC file
 - WES QC file with Picard metrics
 - Germline-Tumor Concordance file
 - Germline Gender Concordance file
 - Clinical Data File (v2.03 schema and normalized tables)
 - MM Spoke-level Clinical Data File

- **Avatar Informatics Tool (Qlik)**
 - Paired molecular and clinical data for cohort identification

ORIEN Avatar: Current Avatar Patients Disease and Clinical Data Summary

Disease	Avatar Patient Count	Core Clinical Data Elements	Spoke Data Elements
BRE - Breast Cancer	482	482	
H&N - Head and Neck Cancer	434	434	
HEM - Multiple Myeloma	393	393	393
SAR - Sarcoma	272	272	
GI - Colorectal Cancer	268	268	
END - Thyroid Cancer	198	198	
GU - Kidney Cancer	197	197	
THO - Lung Cancer	194	194	
GI - Pancreatic Cancer	168	168	
GYN - Ovarian Cancer	166	166	
GU - Prostate Cancer	160	160	
GU - Bladder Cancer	84	84	
GI - Other GI Cancer	71	71	
HEM - Other HEME Disease	67	67	60
GYN - Endometrial Cancer	42	42	
GYN - Other GYN Cancer	42	42	
GI - Gastric Cancer	37	37	
GI - Liver Cancer	37	37	
CUT - Melanoma	25	25	
GI - Esophageal Cancer	22	22	
HEM - Leukemia	20	20	
NEU - Brain Cancer	14	14	
GU - Other GU Cancer	12	12	
OTH - Other Cancer, NOS	11	11	
HEM - Lymphoma	8	8	
THO - Other Thoracic Cancer	3	3	
CUT - Other Cutaneous Cancer	2	2	
OTH - Other Non-Tumor	2	2	
GU - Bladder Cancer and GI - Other GI Cancer	1	1	
HEM - Leukemia and GI - Esophageal Cancer	1	1	
GU - Prostate Cancer and GI - Pancreatic Cancer	1	1	
Total Avatar Patients	3434	3434	453

ORIEN Avatar: Molecular Data QC Parameters

Primary and Secondary QC parameters used to evaluate data generated by HudsonAlpha

QC	Metric	Pass	Flag	Fail
Primary	RNA-Seq			
	PF_Bases (Paired Reads)	>= 20E6 reads		< 20E6 reads
	PCT_Aligned	>= 65%	< 65%	
	PCT_Usable_Solid	>= 55%	< 55%	
	PCT_Usable_MM	>= 35%	< 35%	
	WES			
	PCT_20X	>= 80%		< 80%
	Mean_Coverage_Tumor	>= 100x		< 100x
	Mean_Coverage_Germline	>= 50x		< 50x
Secondary	Concordance			
	SNP Call Rate	> 20,000		< 20,000
	Germline-tumor concordance	>=80.0%		<80.0%
	Germline gender concordance	0 mismatch	>=1 mismatch	

Previous Germline-tumor concordance Metrics:
 8/15/17
 Pass: $\geq 73.0\%$
 Fail: $< 73.0\%$

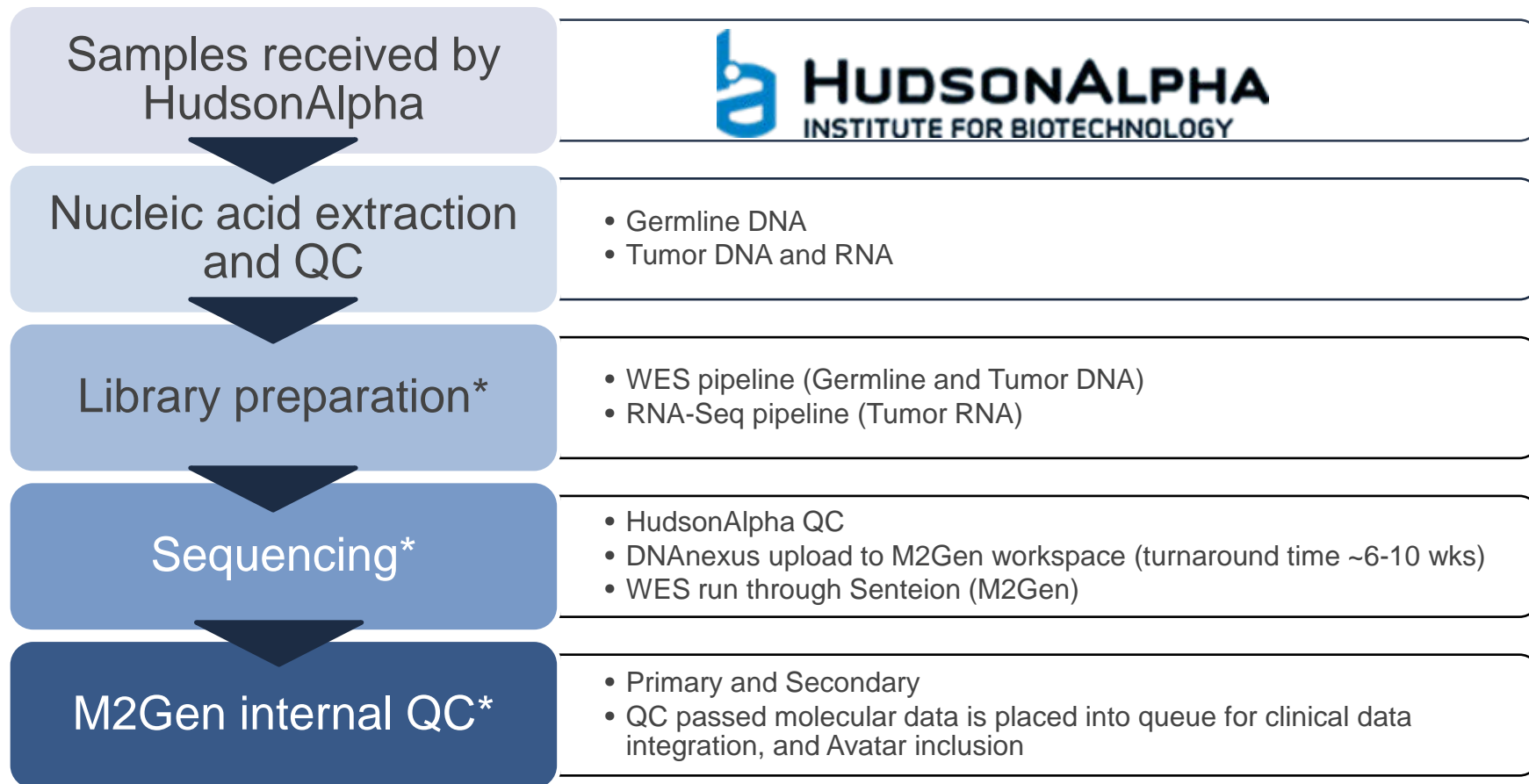
Data fully meeting all QC criteria to be uploaded. Previously released WES pairs “flagged” for low germline-tumor concordance (<73%) remain in Avatar, while new WES data generated with <20000 SNPs or germline-tumor concordance less than 80% are not released.

Data failing to meet basic coverage/reads criteria not uploaded

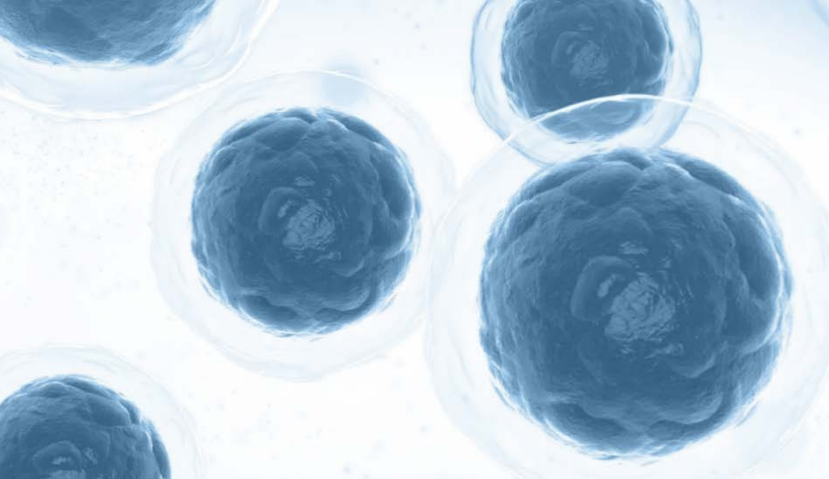
ORIEN Avatar: Definition of Terms

Term	Definition
Avatar Patient	A patient with clinical data, and at least one form of Tumor molecular data (WES Tumor or RNA-Seq Tumor). Patient data is available via sFTP or S3 bucket transfer (clinical data), DNAnexus (molecular data) and Informatics Modules (both clinical and molecular data).
Core Clinical Data Elements	Data elements included in Clinical Data File (v2.03 schema and normalized tables) including diagnosis, treatment and other clinical information.
Spoke Level Clinical Data	Additional clinical data relevant to specific diseases, e.g. multiple myeloma.

ORIEN Avatar: Molecular Data Generation Process



*Degree of attrition at each step is dependent on quantity and quality of nucleic acid extracted/provided



M2GEN

Contact Information

- Hongyue Dai, Ph.D.
206-251-9958
Hongyue.Dai@M2Gen.com
- Odalys Capote
813-745-2296
Odalys.Capote@M2Gen.com
- Page Lane
813-745-2482
Page.Lane@M2Gen.com
- Dan Rehn
813-745-6310
Daniel.Rehn@M2Gen.com
- Janni Mirosevich, Ph.D.
813-384-5032
Janni.Mirosevich@M2Gen.com