AD_____

Award Number: W81XWH-11-1-0488

TITLE: "Development of Advanced Technologies for Complete Genomic and Proteomic Characterization of Quantized Human Tumor Cells"

PRINCIPAL INVESTIGATOR: Dr. Charles Cobbs

CONTRACTING ORGANIZATION: Swedish Health Services

Seattle, WA 98122-4379

REPORT DATE: July 2014

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
July 2014	Annual	15 Jun 2013 – 14 Jun 2014

4. TITLE AND SUBTITLE

"Development of Advanced Technologies for Complete Genomic and Proteomic Characterization of Quantized Human Tumor Cells"

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-11-1-0488

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Dr. Charles Cobbs

Email: charles.cobbs@swedish.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Swedish Health Services
Seattle, WA 98122-4379

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT With the establishment of Glioblastoma (GBM) cell lines from GBM patient's tumor samples and quantized cell populations of each of the parental GBM cell lines, we have completed most of our major aims of this project. We will continue in our efforts in the development and analyses of these quantized cells and develop the genomic and proteomic technology to interpret specific transcriptome and proteome signatures. Whole genome sequencing from two families of GBM patients are now well established and from the basis of the molecular characterization of the tumor development and signatures presented by these tumors for the development of diagnostic signature panels. We have established efficient secretome cell culture conditions to enable proteomic analysis of these quantized cell sub populations and have begun to assemble the protein signatures of GBM tumors underpinned by the comprehensive molecular characterization of these tumors through family WGS, transcriptomics of single cells and secreted protein analysis. This program will ultimately provide new technical capabilities through broad based molecular characterization at the genomic and proteomic analysis of GBM patients and their immediate families and has already provide significant results. This final effort in the molecular characterization of a second patient and their family and the establishment and deployment of a proteomic signature for GBM diagnostics will be completed within the next 12 months.

15. SUBJECT TERMS-
Human cohorts, Glioblastoma, Genomic, Proteomic, Single-cell technologies, Hypothesis-driven, integrative systems approach, Early diagnosis,Patient stratification,Blood protein biomarkers,Quantized cell populations

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	20	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

INTRODUCTION

This collaborative research program will provide important insights into human cancer mechanisms. In particular, we have developed quantitative tools with direct applications for patients with glioblastoma multiforme (GBM), the most common and most aggressive brain tumor in adults. Despite treatment strategies that include gross total resection, post-operative radiation and chemotherapy, most patients develop tumor recurrence within months. Cancer stem cells have been implicated as the presumed cause of tumor recurrence and resistance to therapy. With this in mind, we will utilize GBM patient-derived cell lines and new technologies for transcriptome, miRNAome, proteome, and single-cell analyses to study quantized GBM cell populations and their role in disease progression. This proposal will significantly advance genomic, proteomic and single-cell technologies. The tools proposed here will be generally applicable to all cancer-based studies, as the nature of the tool development is designed to identify and quantify DNA, RNAs, proteins and cells, challenges ubiquitous to all human disease systems.

The expected outcomes of this research program include: 1) a deeper understanding of human GBM disease mechanisms; 2) blood protein biomarkers for use in early diagnosis, assessment of GBM progression, evaluation of drug treatment effectiveness, and early detection of disease recurrence; 3) new strategies for advanced genomic sequencing of quantized cancer cells and their normal counterparts to identify cancer-driver mutations; 4) new technologies for transcriptome, miRNAome, proteome, and single-cell analyses, and 5) the creation of quantized GBM cell lines that can be used for general molecular characterization as well as to assess GBM biology and the effectiveness of existing drugs in reacting with these cell types. To achieve these goals we have pursued the following aims:

Specific Aim 1. Isolate up to 1000 cells from each of five human GBMs and quantify initially 500 different transcripts from each cell (transcription factors, CD molecules, relevant signal transduction pathways, etc). Determine whether computational analyses can classify these cells into discrete quantized cell types.

Specific Aim 2. Sort the disassociated tumor cells from GBM into their quantized cell populations using cell sorting/CD antibodies to each quantized cell type for functional analyses and establish primary cell lines. These will be used for molecular analyses at the genome, transcriptome, miRNAome and selected proteome levels.

Specific Aim 3. Assess 20-40 candidate blood biomarkers in the bloods of 100 GBM patients with regard to their ability to stratify disease, assess disease progression and predict at an early stage GBM recurrence. Eventually we will use these biomarkers to assess the effectiveness of therapy.

Specific Aim 4. Ten to 20 cells from each major quantized GBM cell type from two patients will be used to determine the complete genome sequences. We will also determine the normal genome sequences of each patient and their family members to enable the Mendelian-based error correction process. The mutations will be analyzed against quantitative changes in the transcriptomes, miRNAomes and proteomes and against the relevant biological networks.

Specific Aim 5. Analyze the quantized cell populations for their responses (transcriptome, miRNAome, etc) to the perturbations of key GBM-relevant molecules (e.g. nodal points in networks) by RNAi perturbations as well as their responses to GBM-relevant drugs and natural ligands.

BODY

Specific Aims 1, 2 & 4. During the past year, we continued to collect GBM tumor samples to establish the patient cohort available for molecular analysis, genome sequencing, and quantitative assays. The Ivy Center for Advanced Brain Tumor Treatment at the Swedish Neuroscience Institute (SNI) has collected tumor tissue eligible for this program from over forty GBM patients to date. Several tissue specimens were processed to establish primary GBM cell lines and tested to confirm the presence of the stem cell phenotype. In brief, tumor samples were treated with Accutase (Sigma) immediately after surgical resection, and single cell suspensions were plated in neurobasal media with epidermal growth factor (EGF) and fibroblast growth factor (FGF-2) as described (*1, 2*). Although the criteria that define cancer stem cells are controversial (*3-6*), it is generally accepted that GBM stem cells must demonstrate self-renewal, multipotency, and tumor-initiating ability *in vivo* (*7*). Our patient derived cultures contain cells capable of forming self-renewing spheres (signifying the presence of a self-renewing stem cell population), are multipotent (Figure 1), and have the ability to form tumors in NOD/SCID mice 6-8 weeks after cell inoculation (Figure 2). These results demonstrate the presence of stem-like cells within our patient-derived GBM

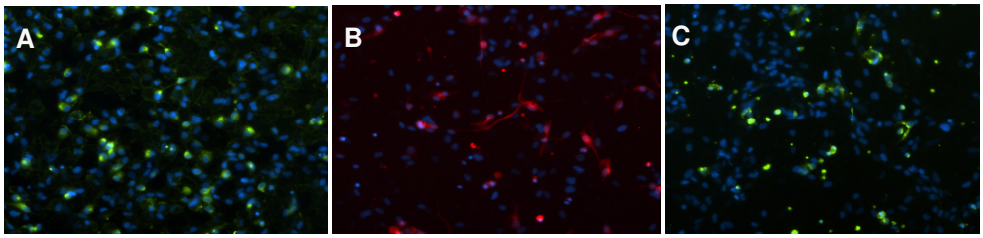


Figure 1. Immunostaining for differentiation markers: (A) GFAP/astrocytes, (B), TUJ-1/neurons and (C) O4/oligodendrocytes. Parental GBM cultures established at SNI are multipotent and able to differentiate into astrocytes, neurons and oligodendrocytes.

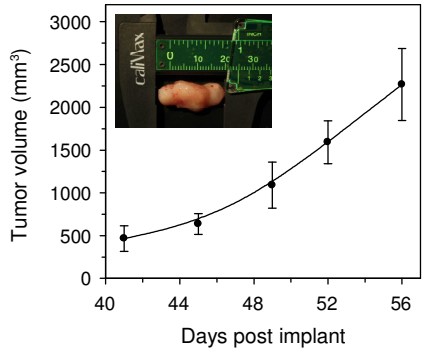


Figure 2. Average tumor volume (n = 5) in immunocompromised mice six weeks after implantation of GBM-patient derived cells (SN186). *Inset*: Tumor mass removed from host. Tumor initiation *in vivo* confirms the presence of stem cell populations within the heterogeneous cell culture.

cultures, and verify that these cell lines are suitable for the proposed research.

SN#	Gender	Age	Histopathology	Resection	Subtype	MGMT	Chemotherapy	Radiation	Survival (days)
143	Male	75	GBM (Gliosarcoma), grade IV	Left Temporal	Mesenchymal	Unmethylated	Not available	Not available	323
186	Male	76	GBM, grade IV	Right Temporal	Proneuronal	Unmethylated	140mg TMZ, over 11 weeks (concurrent with radiation).	IMRT, 4500 cGy in 25 fractions, over 6 weeks.	459
243	Male	57	GBM, grade IV	Right Frontal	Proliferative	Methylated	160mg TMZ, concurrent, 6 weeks; 400mg TMZ, maintenance 5x/mo, 38 weeks; 160mg TMZ, maintenance 21x/mo, 8 weeks; 400mg TMZ, maintenance 5x/mo, 8 weeks.	IMRT, 4140 cGy in 23 fractions, concurrent, 3 weeks. IMRT, 1800 cGy in 10 fractions, boost, 3 weeks.	Alive
291	Female	63	GBM, grade IV	Right Parietal	Mesenchymal	Methylated	150mg TMZ, every 2 weeks 5 days cycle, for 54 weeks.	IMRT, over 8 weeks. Stereotactic, 2500 cGy in 5 factions, 1 week.	Alive
348	Female	49	GBM, grade IV	Right Frontal	Not determined	Unmethylated	105mg TMZ, concurrent, 7 weeks.	IMRT, 5940 cGy in 33 fractions, 7 weeks.	123

Table 1. Clinical diagnosis, treatment history and survival of GBM patients used in this study.

We have transferred several of these GBM-derived cultures (as summarized in Table 1) to our collaborators at the Institute for Systems Biology (ISB; Award Number W81XWH-11-1-0487, Dr. Robert Moritz) for the generation of quantized cell populations. These cell populations serve as the foundation for subsequent genomic, transcriptomic, and proteomic studies. In particular, the GBM specimens SN243 and SN291, for which we have consenting family members, were selected for complete molecular analyses. We collected blood (processed as plasma and peripheral blood mononuclear cells [PBMCs]), from both SN243 and SN291 patients, and their family members during the last year (as summarized in Table 2). This completes the specimen cohort required for moleucular analyses at the genome, transcriptome, miRNAome and proteome levels (Specific Aims 1 and 2).

Patient #	Family #	Relationship	Gender	Age
SN243 Male, 56	SN243-P1	Parent	Male	89
	SN243-P2	Parent	Female	86
	SN243-C1	Child	Female	35
SN291 Female, 62	SN291-S1	Sibling	Female	73
	SN291-C1	Child	Male	35
	SN291-C2	Child	Female	41
SN348 Female, 48	SN348-P1	Parent	Male	74
	SN348-P2	Parent	Female	75
	SN348-S1	Sibling	Female	46
	SN348-C1	Child	Female	26

Table 2. Blood samples collected for whole genome family sequencing

A number of quantized cell populations have been successfully established from the corresponding SN243 and SN291 parental cell lines (Figure 3). To generate quantized cell populations a single cell clonal culture technique, integrated with single cell sorting using the BD FACS Aria II, was developed. Approximately 60% of the sorted cells formed colonies (>100 cells) and were collected and frozen for further analysis. For each primary tumor line, we established clonal cultures which exhibited distinct morphological phenotypes. Given that each clone presumably carries a uniform genome, it is suitable for whole genome sequencing. 5 clones have been selected from each patient for subsequent 'omic analysis.

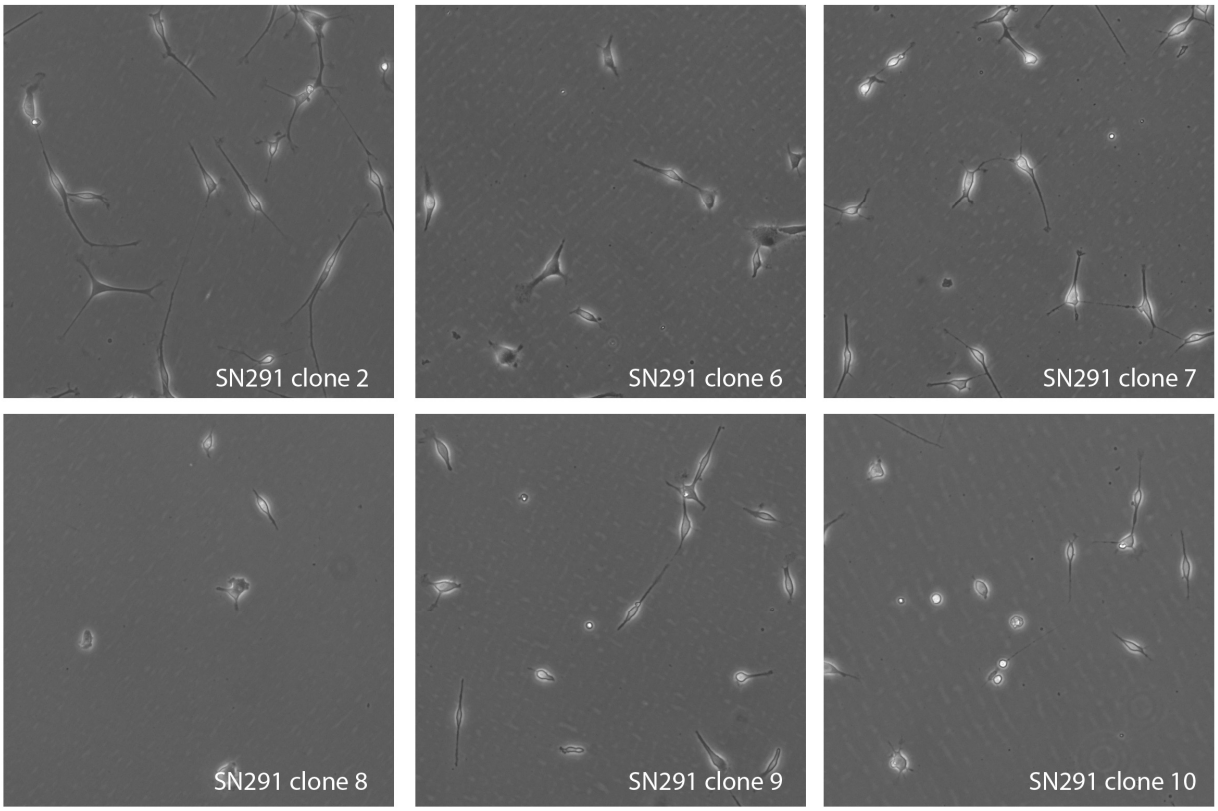


Figure 3. Establishment of single cell clonal cultures from GBM patient SN291. A total of 12 clonal cultures have been generated from this patient sample. In brief, cells were cultured on plates coated with laminin and grown under serum-free conditions with stem cell media supplement with B27, N2 and the growth factors EGF and bFGF. Single cell sorting was performed using the BD FACS Aria II. Approximately 60% of the sorted cells formed colonies (>100 cells) and were used for further analysis.

Sample preparation for whole genome sequencing

Genomic DNA has been extracted from both SN243 and SN291 specimens for whole genome sequencing. For SN291, the specimens include tumor tissue, normal sequence coverage of primary tumor cell culture, 5 quantized cell populations, patient-derived PBMCs and PBMCs from the patient's two children. The samples for SN243 include tumor tissue, primary tumor cell culture, 5 subclones, patient-derived PBMCs, and PBMCs from both the patient's parents and one child (Figure 4). The quality of each DNA prep was validated prior to submission to Complete Genomics (Mountain View, CA) for whole genome sequencing analysis.

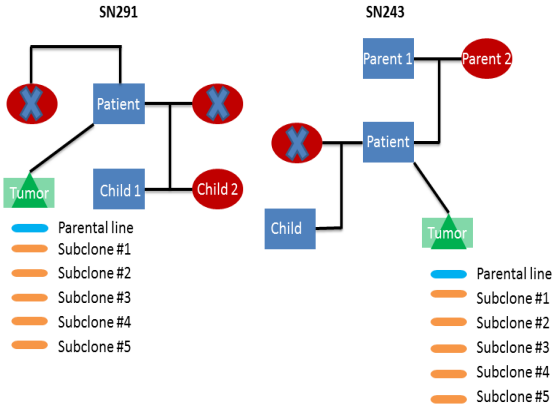


Figure 4. Patient selection and family history of samples submitted for whole genome sequencing.

A total of 22 DNA samples were submitted to Complete Genomics. All ten DNA samples from patient SN291 and family members passed quality control by CGI. High quality whole genome sequencing data was generated and sent back to our collaborators at ISB for further analysis (Figure 5). As shown in Table 2, eleven out of the twelve SN243 samples passed initial CGI quality control, and were moved down their pipeline for sequencing analysis. However, we were recently informed that DNA samples from two clonal populations (4 and 12) unfortunately failed insertion of one of the adapters. SN243 clone 4 and clone 12 are therefore being expanded again. Once sufficient cells are available, genomic DNA will be extracted from SN243 clone 4 and clone 12 for resubmission to CGI.

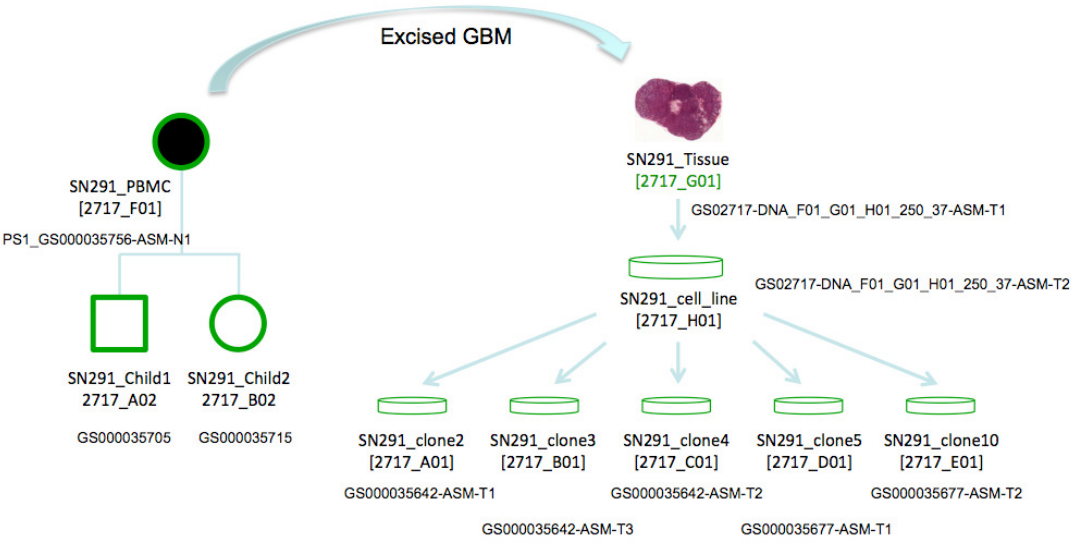


Figure 5. Overview of patient SN291 genome dataset. For each sample, the descriptive identifier, the vendor sample identifier between square brackets, and the vendor assembly identifier is shown.

Sample ID	Customer Sample ID	Customer Subject ID	Tumor Status	QC Status	Details	DNA Vol. Reported by Customer (µl)	DNA Vol. Measured by CGI (µl)	DNA Conc. Reported by Customer (ng/µl)	DNA Conc. Measured by CGI (ng/µl)	Available DNA (µg)	Gender Reported	Count of ChrY SNPs	Total Count of Called SNPs	Gender Match to Reported Gender
GS02741-DNA_A01	SN243 PBMC	PBMC	Non-Tumor	Passed		190	182	78.8	77.8	13.8	Male	9	87	Match
GS02741-DNA_B01	SN243 TISSUE	Tissue	Tumor	Passed		100	98	145.3	167.9	14.8	Male	9	91	Match
GS02741-DNA_C01	SN243 P1	parent 1	Non-Tumor	Passed		70	53	212.3	253.9	13.5	Male	9	91	Match
GS02741-DNA_D01	SN243 P2	parent 2	Non-Tumor	Passed		88	84	170.3	171.7	11.0	Female	0	78	Match
GS02741-DNA_E01	SN243 C1	child1	Non-Tumor	Passed		200	174	81.8	85.8	11.4	Female	0	88	Match
GS02741-DNA_F01	SN243 PARENTAL	parental cell	Tumor	Passed		50	46	200.3	140.8	6.4	Male	9	86	Match
GS02741-DNA_G01	SN243 CLONE 1	clone 1	Tumor	Failed	Quantity Failed	50	43	298.3	56.9	2.4	Male	9	82	Match
GS02741-DNA_H01	SN243 CLONE 2	clone 2	Tumor	Passed		50	44	100.3	87.7	3.8	Male	9	90	Match
GS02741-DNA_A02	SN243 CLONE 4	clone 4	Tumor	Passed		50	44	85.3	82.1	4.0	Male	9	87	Match
GS02741-DNA_B02	SN243 CLONE 6	clone 6	Tumor	Passed		180	169	53.1	45.1	7.6	Male	9	82	Match
GS02741-DNA_C02	SN243 CLONE 7	clone 7	Tumor	Passed		50	43	130.4	200.0	8.6	Male	6	66	Match
GS02741-DNA_D02	SN243 CLONE 12	clone 12	Tumor	Passed		90	78	70.4	54.1	4.2	Male	9	79	Match

Table 3. CGI quality control template for 12 genomic DNA samples from SN243 family.

Analysis of Whole genome data: Karyotype computed from genome data

Our collaborators at ISB have developed a sophisticated method for the identification of aneuploidies at high resolution. This has been applied in the analysis of the SN291 genomes (see PI Moritz report). Results are presented in Figure 6 below.

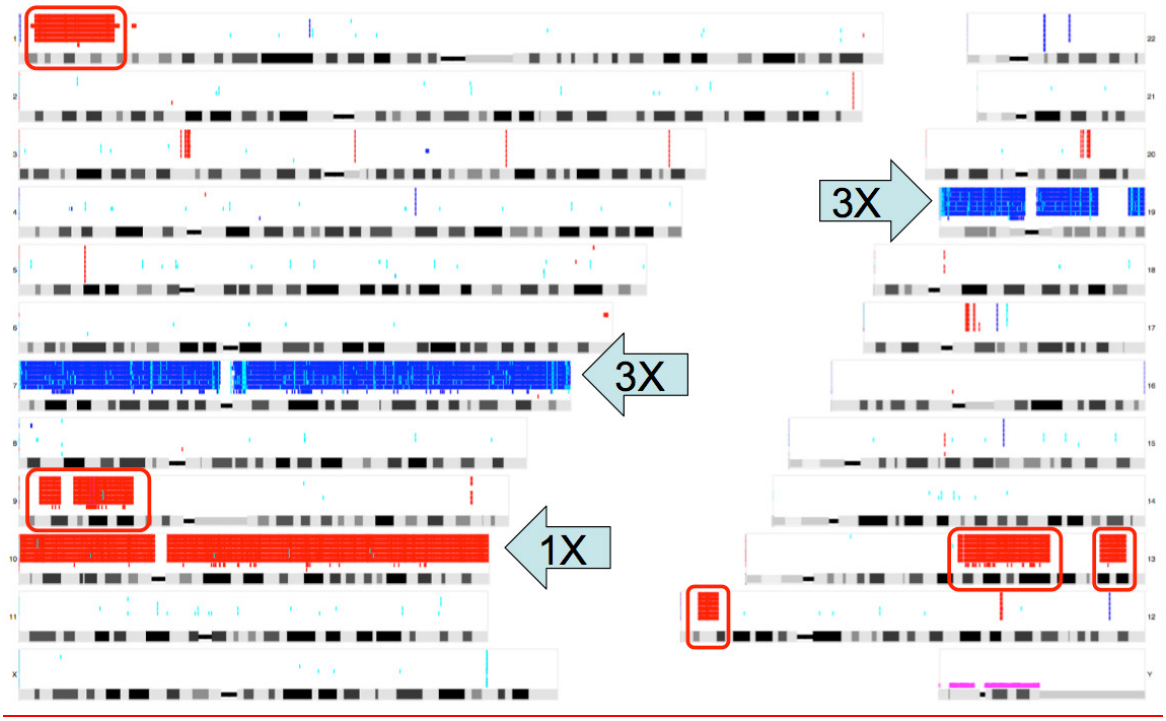


Figure 6. Computed karyotype. For each chromosome, we depict (bottom to top) the computed copy numbers observed for SN291's PBMC genome, cancer tissue, cell line and subclones. Red denotes deletions (haploid), blue represents expansions (triploid, with light blue representing tetraploid or higher).

The complete loss of one copy of chr10 was observed in the SN291 cell line and corresponding subclones. The fragmented signal for the cancer tissue indicates the presence of this aneuploidy in a subpopulation of the mixed tissue. The chromosomal deletion is not observed in the PBMC sample, as expected. Similarly, we observed large-scale but partial losses in chromosomes 1, 9, 12 and 13. Conversely, we observed an extra copy (triploidy) of chr7 and most of chr19.

Relationship between subclones of SN291

Based on the aneuploidy analysis, it is evident that the five subclones are independent of each other (Figure 7). Each subclone presents a small number of minor private aneuploidies, none of which is shared by two or more subclones.

Variant analysis

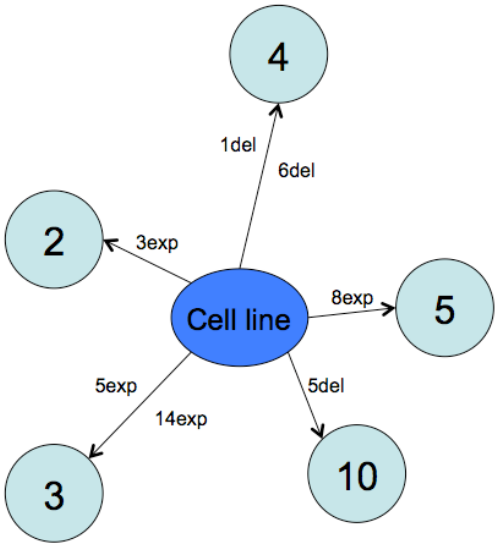
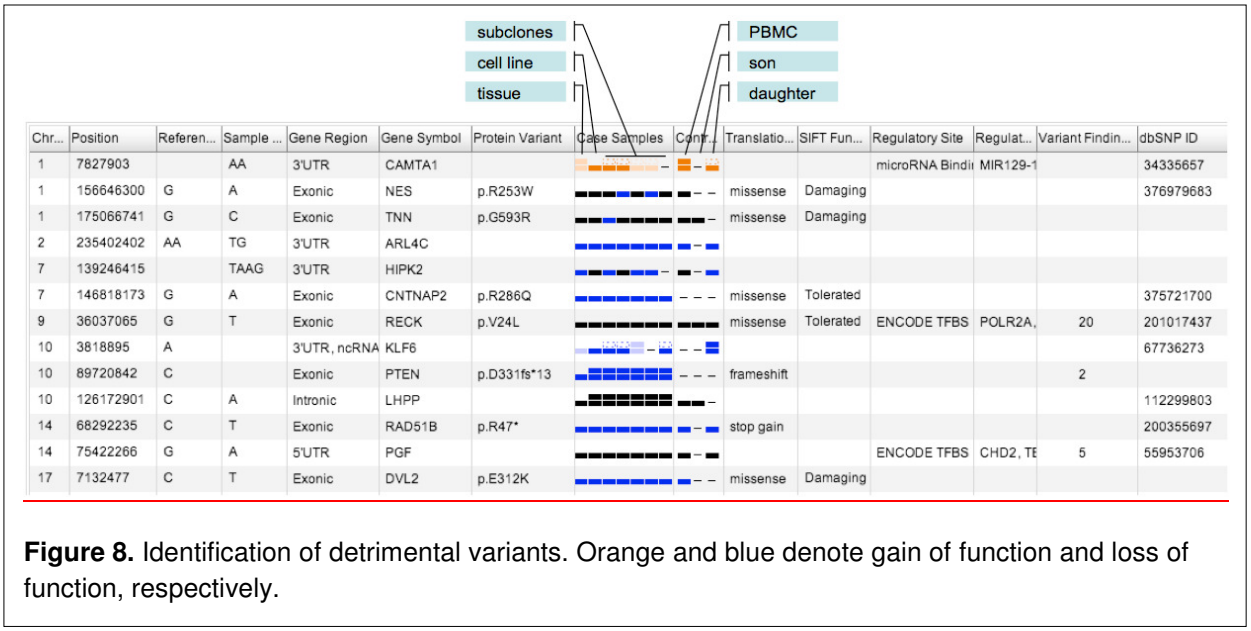


Figure 7. Relationships between the SN291 subclones and the parental cell line. Each subclone presents a very small number of private aneuploidies (deletions or expansions).

Ingenuity Variant Analysis is a web-based application that helps researchers study human disease by identifying causal variants from human sequencing data. Ingenuity Variant Analysis was applied to the ten genome sequences to identify candidate variants associated with the glioblastoma phenotype, using the tissue, cell line and five subclones as "cases". We required candidate variants to be predicted deleterious, observed in at least three "cases" with quality >= 35, and with population frequency under 1%. We used Ingenuity's knowledgebase to select cancer driver variants directly affecting genes known to be involved in GBM. As shown in Figure 8, a number of interesting gene mutation candidates were identified. Of particular interest is a stop-gain SNV in the RAD51B gene, present in heterozygous form in the genome of the patient (PBMC), the cancer tissue, the cell line and all the subclones. This variant is very rare, with a population frequency of 0.0079% (as computed using Kaviar genome database) and is confirmed by its presence in the daughter (but absent in the son). A second variant of interest is a novel missense SNV in DVL2, predicted to be deleterious.



Chr...	Position	Referen...	Sample ...	Gene Region	Gene Symbol	Protein Variant	Case Samples	Contr...	Translatio...	SIFT Fun...	Regulatory Site	Regulat...	Variant Findin...	dbSNP ID
1	7827903		AA	3'UTR	CAMTA1						microRNA Bindi	MIR129-1		34335657
1	156646300	G	A	Exonic	NES	p.R253W			missense	Damaging				376979683
1	175066741	G	C	Exonic	TNN	p.G593R			missense	Damaging				
2	235402402	AA	TG	3'UTR	ARL4C									
7	139246415		TAAG	3'UTR	HIPK2									
7	146818173	G	A	Exonic	CNTNAP2	p.R286Q			missense	Tolerated				375721700
9	36037065	G	T	Exonic	RECK	p.V24L			missense	Tolerated	ENCODE TFBS	POLR2A,	20	201017437
10	3818895	A		3'UTR, ncRNA	KLF6									67736273
10	89720842	C		Exonic	PTEN	p.D331fs*13			frameshift				2	
10	126172901	C	A	Intronic	LHPP									112299803
14	68292235	C	T	Exonic	RAD51B	p.R47*			stop gain					200355697
14	75422266	G	A	5'UTR	PGF						ENCODE TFBS	CHD2, TE	5	55953706
17	7132477	C	T	Exonic	DVL2	p.E312K			missense	Damaging				

Figure 8. Identification of detrimental variants. Orange and blue denote gain of function and loss of function, respectively.

RNA-Seq Data Analysis

A number of published computational algorithms have been evaluated for analyzing RNA-seq data (Figure 9). To fully extract gene expression information from glioblastoma tumor tissues and clonal populations for biomarker discovery, we have decided to use TopHat based on its more general application, robust performance, and wide acceptance in the community, in addition to the fact that our single cell RNA-seq data is of lower coverage.

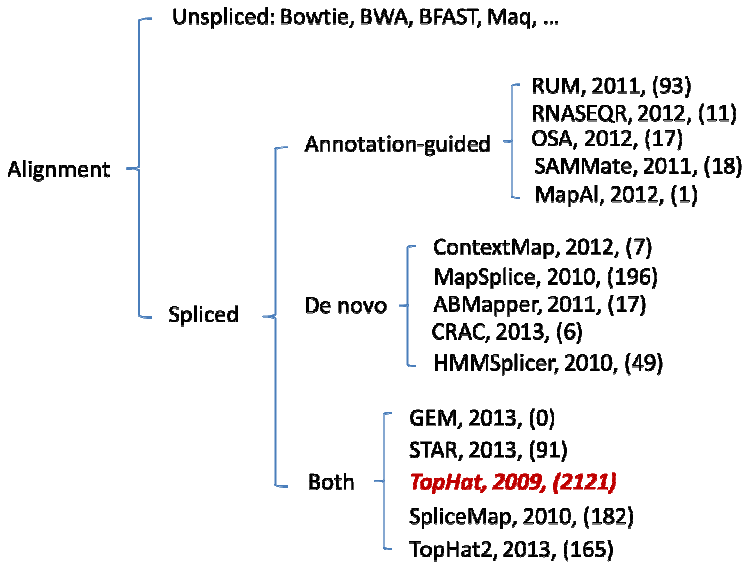


Figure 9. Evaluation of published RNA-seq data analysis algorithms. Numbers in parentheses indicate number of citations for each program.

Single Cell RNA-Seq data for cultured SN291 tumor cells.

We have successfully generated single cell RNA-Seq data from SN291 tumor cells. cDNA libraries from 96 individual single cells, derived from patient SN291, were prepared, indexed and subjected to next generation sequencing on an Illumina HiSeq platform. Approximately 1-2 million quality reads from 87 cells were generated; ~60% of reads can be mapped concordantly to the human genome. For nine cells insufficient sequencing reads were produced and thus these were excluded from further analysis (Table 4).

Principle component analysis and network mapping.

Principle component analysis was performed on the 87 single cell transcriptomes for SN291. As shown in Figure 10, several distinct cell clusters were identified. Our collaborators at ISB used their previously published work (PNAS, 2011) to evaluate the enrichment pattern for CD133+ gene signatures. SN291 cells bearing CD133+ signature (red) show distinct separation from those cells negative for the signature. One cell (purple) shows a strong enrichment for Wnt signaling pathway genes.

Sample	AlignedPairs	ConcordantPairs	DisconcordantPairs	C1 type	Sample	AlignedPairs	ConcordantPairs	DisconcordantPairs	C1 type
g01	899,235	49.7%	34.0%	1	g49	1,326,584	66.1%	19.7%	1
g02	1,191,278	47.5%	36.3%	1	g50	1,709,060	63.1%	21.8%	1
g03	1,612,721	62.1%	20.9%	1	g51	1,161,220	65.7%	21.2%	1
g04	615	39.5%	43.1%	1	g52	874,909	47.9%	35.6%	1
g05	1,647,110	59.9%	23.6%	1	g53	1,119,424	55.0%	28.8%	1
g06	1,797,059	60.4%	22.1%	1	g54	1,211,442	51.8%	33.1%	1
g07	1,686,501	66.0%	18.6%	1	g55	1,215,567	53.4%	30.6%	1
g08	1,351,292	55.6%	23.4%	2	g56	1,142,071	53.9%	29.1%	1
g09	1,605,567	61.6%	21.5%	1	g57	1,324,289	55.3%	28.1%	1
g10	1,019,528	49.0%	33.4%	2	g58	1,711,576	66.0%	17.7%	1
g11	1,176,268	49.4%	34.2%	1	g59	1,426,244	64.9%	20.5%	1
g12	1,079,971	57.7%	26.6%	1	g60	1,330,170	65.6%	18.6%	1
g13	1,572,060	61.4%	24.6%	1	g61	1,706,541	64.0%	22.5%	1
g14	1,009,804	56.1%	28.1%	1	g62	1,682,185	64.9%	20.4%	1
g15	1,161,573	57.2%	26.6%	2	g63	1,372,796	56.0%	25.8%	1
g16	1,101,282	56.9%	25.7%	1	g64	1,291,539	66.3%	16.3%	1
g17	1,040,458	71.5%	13.6%	1	g65	1,006,918	60.5%	23.2%	1
g18	1,912,866	64.7%	18.3%	2	g66	546,529	50.9%	33.4%	1
g19	1,197,388	57.9%	26.1%	3	g67	1,292,106	63.8%	21.1%	1
g20	1,187,574	55.1%	28.0%	1	g68	211,436	79.7%	8.1%	1
g21	1,611,582	53.7%	26.5%	1	g69	1,306,388	55.0%	28.9%	1
g22	1,419,808	58.4%	24.9%	1	g70	922,079	47.6%	36.3%	1
g23	1,348,420	56.6%	26.6%	1	g71	1,107,861	51.5%	32.4%	1
g24	787,207	59.6%	24.3%	1	g72	257,779	74.8%	10.3%	2
g25	1,460,197	65.5%	19.9%	1	g73	701,982	44.3%	39.4%	1
g26	2,016,625	66.0%	20.3%	1	g74	1,152,883	62.6%	22.9%	1
g27	1,968,362	62.2%	23.0%	1	g75	1,355,419	69.1%	19.5%	1
g28	1,138,651	57.9%	23.0%	1	g76	914,087	63.1%	21.6%	1
g29	1,312,753	57.2%	27.8%	1	g77	1,418,660	70.3%	16.3%	1
g30	1,631,678	63.2%	21.4%	1	g78	1,250,458	58.7%	26.5%	1
g31	1,472,455	65.8%	18.2%	1	g79	1,669,264	65.9%	18.4%	1
g32	1,974	77.0%	15.1%	0	g80	1,036	80.6%	11.1%	0
g33	1,429,980	61.6%	23.6%	1	g81	1,304,525	59.4%	25.0%	1
g34	1,125,710	55.4%	30.7%	1	g82	1,708,550	62.9%	21.7%	1
g35	999,367	52.8%	32.0%	1	g83	1,493,367	68.9%	18.1%	2
g36	1,244,461	56.0%	25.4%	1	g84	508,605	56.4%	26.8%	1
g37	1,619,458	62.1%	22.5%	1	g85	279,937	71.8%	10.0%	1
g38	3,827,684	74.7%	11.5%	1	g86	731,840	47.7%	35.7%	1
g39	2,130,139	67.8%	15.8%	1	g87	980,536	51.1%	33.0%	1
g40	19,340	78.7%	7.4%	0	g88	1,585	82.6%	9.7%	0
g41	1,741,414	68.7%	17.1%	1	g89	1,145,143	47.6%	35.0%	1
g42	1,482,177	68.8%	18.0%	1	g90	1,706,828	65.2%	18.0%	1
g43	1,283,590	59.6%	25.3%	1	g91	1,439,922	65.7%	20.4%	2
g44	1,175,265	55.7%	27.6%	1	g92	1,196,507	58.4%	22.5%	1
g45	1,643,689	63.7%	18.8%	1	g93	1,605,522	66.5%	16.1%	1
g46	2,080,139	65.8%	16.6%	1	g94	1,698,874	71.6%	13.1%	1
g47	1,203,681	52.6%	31.8%	1	g95	1,925,914	62.6%	21.3%	1
g48	1,243,135	56.4%	28.9%	1	g96	1,117	83.6%	10.2%	0

Table 4. RNA-seq analysis of 96 single cells from patient SN291.

Refinement of RNA samples from clonal populations for RNA-seq analysis

One of the key components of this study will be to compare gene expression profiles to the whole genome sequencing data generated from the same clonal populations. In order to achieve this goal, total RNAs have been extracted from the parental SN243 and SN291 cell lines and five to six of their subclones. RNA integrity has been validated and sample prep for the RNA-seq will be completed in the coming months prior to sample submission.

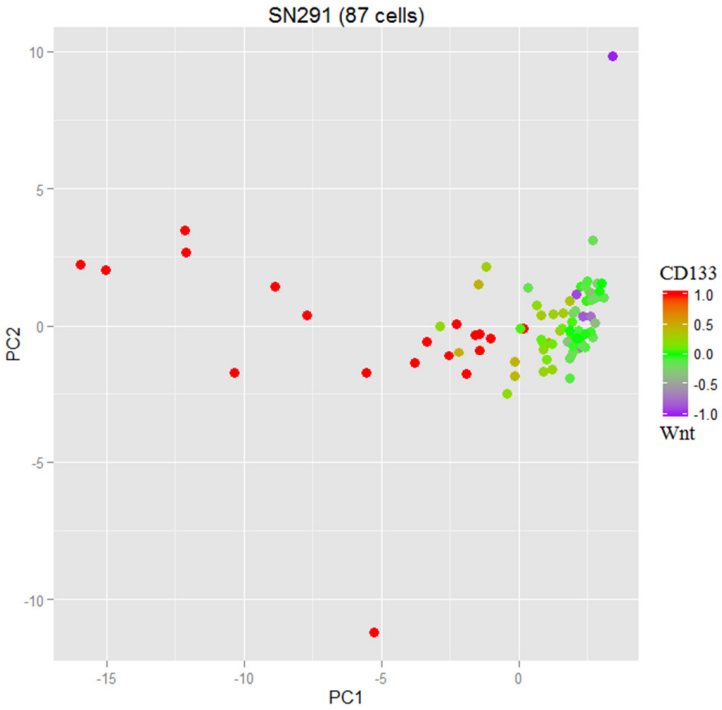


Figure 10. Principle component analysis of RNA-seq data generated from individual cells from the SN291 parental culture. Each dot represents a single cell. Color gradient indicates enrichment score for either our published CD133+ gene signature (PNAS, 2011) or Wnt pathway genes.

Specific Aim 3. Parental tumor cell lines and subclones have been expanded for proteomic analysis aimed at identifying candidate protein biomarkers for glioblastoma. Extraneous proteins were removed from culture conditions to allow for the identification of proteins secreted directly by quantized cell populations. Over the last year, cell pellets and secreted protein preparations have been collected for the SN291 parental line and five subclones. Similarly, cell pellets and secreted protein preparations have been prepared for the SN243 parental line and six subclones. SN260 and SN348 parental cell lines have also been prepped for comparative proteomic analysis. In summary, cell pellets and secreted protein samples were prepared from four parental cell lines and eleven subclones derived (Figure 11).

We have begun deep proteome analysis of the secretomes by performing off-gel fractionation of the peptide digests of each of the secretome preparations and have combined the analysis of each of the fractions to prepare lists of differentially secreted proteins from each of the cell lines (progress to date shown in Table 5). The cell pellets have been further processed by our collaborators at ISB to study the cell surface proteome using N-glyco capture technology and subsequent analysis by liquid chromatography-mass spectrometry (LC-MS/MS). The N-glycosylated proteins are of interest in the context of biomarker identification strategies and

expected to provide insight in differentially expressed signatures of glioblastoma patients. The secretome preparations of the four parental cell lines (SN243, SN291, SN260 and SN348) and eleven subclones yielded enough material for subsequent LC-MS/MS analysis. The 15 samples were subjected to tryptic digestion using a standard protocol including the reduction of disulfide bonds with dithiothreitol and alkylation of the sample with iodoacetamide. Samples were analyzed in duplicate using a high resolution QExactive mass spectrometer (Thermo Fisher Scientific) allowing peptide fragmentation by higher-energy collisional dissociation. Peptides were separated on a reversed phase column using a particular long gradient of 4h and nano-LC conditions to allow highly sensitive in depth analysis of the secretome of each tumor cell sample.

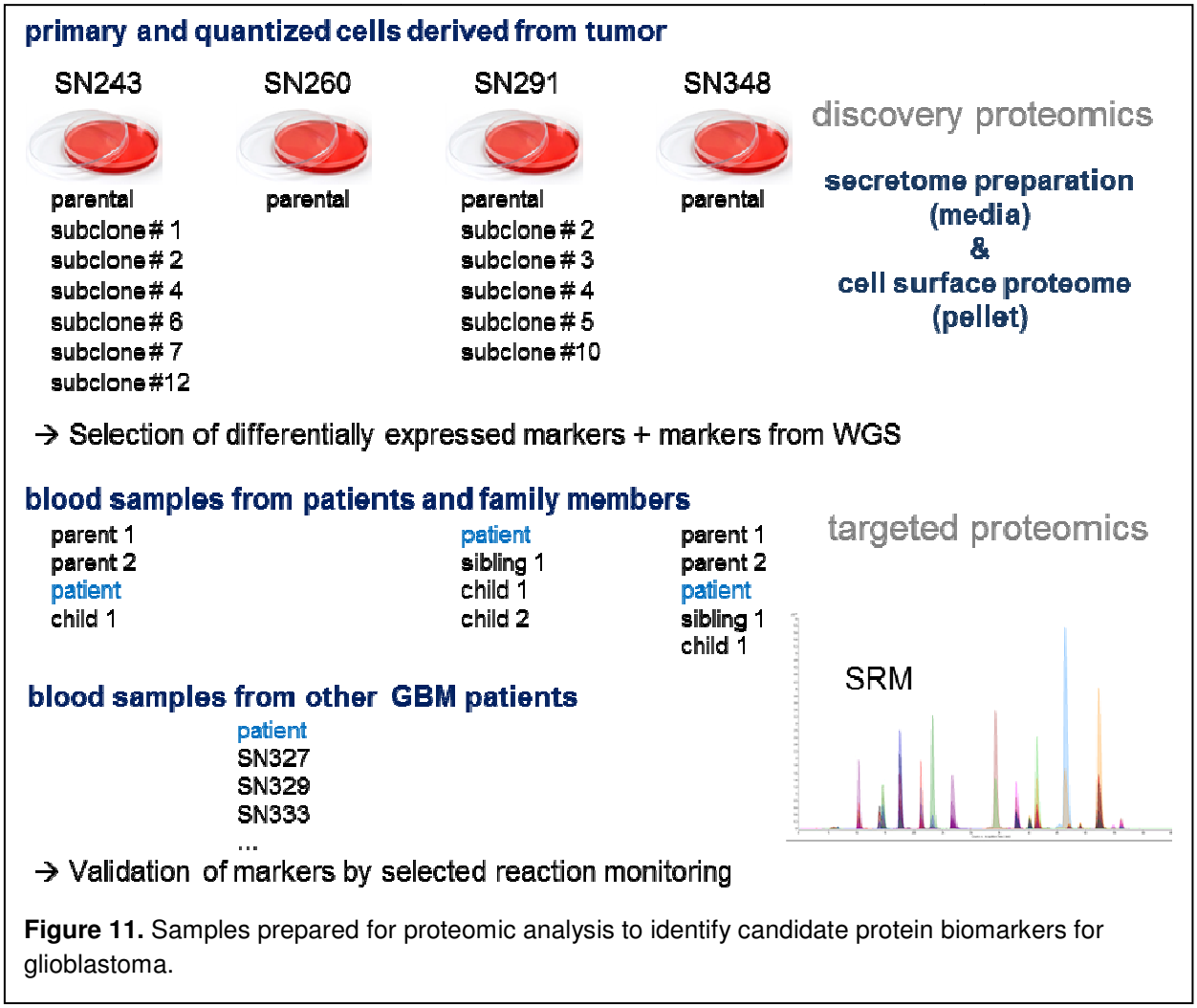


Figure 11. Samples prepared for proteomic analysis to identify candidate protein biomarkers for glioblastoma.

Proteomic data are analyzed through sequence database searching using the software tool suite of the Trans-Proteomic-Pipeline (developed at ISB) for the correct assignment of MS spectra to peptides and to infer the proteins from these peptide identifications. A standard database would allow detection of known proteins but not the detection of mutational changes

from the tumor genome or the tumor derived quantized cells. To include such mutations, an extended cancer genome specific database will be generated that considers the results from the whole genome sequence and allows for a correlation of specific mutations arising from these tumor cells on the proteome level. This will be done for all SN291 samples (for which full genome sequence information is available) and subsequently for SN243 once whole genome sequencing results are received from Complete Genomics.

Proteomics - Secretome Analysis of Parental Cell Lines and Subclones							
cell line	**type**	**cell culture**	**sample preparation**	**digestion**	**LC-MS/MS analysis**	**data analysis**	**comment**
SN291	parental	✓	✓	✓	✓	in progress	
media blank 1		✓	✓	✓	✓	in progress	
SN291_noS	parental	✓	✓	✓	✓	in progress	supplementals further removed
SN291	subclone 2	✓	✓	✓	✓	in progress	
SN291	subclone 3	✓	✓	✓	✓	in progress	
SN291	subclone 4	✓	✓	✓	✓	in progress	
SN291	subclone 5	✓	✓	✓	✓	in progress	very slow growing
SN291	subclone 10	✓	✓	✓	✓	in progress	
SN243	parental	✓	✓	✓	✓	in progress	
media blank 2		✓	✓	✓	✓	in progress	
SN243_noS	parental	✓	✓	✓	✓		supplementals further removed
media blank 3		✓	✓	✓	✓		
SN243	subclone 1	✓	✓	✓	✓	in progress	
(green = subclones selected for proteomics)	subclone 2	✓	✓	✓	✓	in progress	
	subclone 3	✓					
	subclone 4	✓	✓	✓	✓	in progress	
	subclone 5	✓					
	subclone 6	✓	✓	✓	✓	in progress	
	subclone 7	✓	✓	✓	✓	in progress	
	subclone 8	✓					
	subclone 9	✓					
	subclone 10	✓					
	subclone 12	✓	✓	✓	✓	in progress	
SN260	parental	✓	✓	✓	✓	in progress	
SN348	parental	✓	✓	✓	✓	in progress	

Table 5. Progress on proteomic secretome analysis of parental cell lines and subclones.

Biomarker candidates derived from this discovery proteomic analysis will be correlated with the data derived from the transcriptome and whole genome analysis to define the final list of candidates that will be subjected to targeted quantitative proteomic selected-reaction monitoring (SRM) analysis. Blood samples have been collected from patient SN243 and SN291, and of

three direct family members from each patient. We have also collected additional blood plasma samples from patient SN348 and four direct family members as well as other patients (not selected for whole genome sequencing) to expand the patient population available for targeted SRM analysis of the selected candidates. This will enable us to evaluate GBM specific tumor markers in this cohort of patients as well as the larger pool of 100 plasma samples from both GBM patients and normal subjects.

Database construction for cancer derived mutational proteome analysis.
Our collaborators at ISB have begun to construct a database for cancer derived mutational proteome analysis using the existing whole genome sequencing data from SN291. This will be expanded to include SN243 once the data is available (see PI Moritz report).

Specific Aim 5. We have developed a drug screening protocol for primary GBM cell lines (*1*). In our previous study, SN143 and SN186 parental cells lines were tested against a library of 2,000 compounds composed primarily of FDA approved compounds (50%), natural products (30%), and other bioactive components (20%) (MicroSource Spectrum collection, MicroSource Discovery Inc.), to identify inhibitors of GBM stem cell (GSC) proliferation. As shown in Table 6, this drug screening assay was able to identify approximately 100 compounds that were cytotoxic against SN143 and SN186. The compounds identified represent multiple classes of drugs and natural products, including antineoplastics, cardiotonics, antihelminthics, and others. For Aim 5, our goal is to apply this methodology to quantized cells and use this procedure to identify therapeutic agents with the potential to target the stem cell population. In particular, we will compare the drug responsiveness of the parental cultures with the corresponding sub clone populations. We hope this will provide valuable information that can be

Class[a]	Total in class	Active agents	
		SN143	SN186
Alcohol antagonist	3	1	1
Antihelminthic	33	7	7
Antiarrhythmic	24	0	1
Antibacterial	227	11	11
Antifungal	55	5	5
Antineoplastic	115	29	28
Antihyperlipidemic	12	3	4
Antihypertensive	63	1	1
Anti-infective	11	2	3
Antipsychotic	22	1	1
Cardiotonic	14	10	10
Diuretic	16	0	0
H1 antihistamine	11	1	0
Immunosuppressant	5	1	0
Psychotropic	9	0	1
Sclerosing agent	2	1	1
Vasodilator	35	0	0
Undetermined activity	444	19	22
Total agents	1101b	92	105

Table 6. Pharmacological classes for inhibitors of GSC proliferation of SN143 and SN186 (*1*).

translated to the clinic and used to design effective treatment strategies for future GBM patients. Given that SN243 and SN291 samples have been used for whole genome sequencing, we will apply a similar drug screening approach to both these primary cell lines and their corresponding subclones. Taking into consideration the number of cells required for drug screening and the growth kinetics of the quantized cells, we have decided to pursue a 160-oncology focused library. This library is composed of FDA approved antineoplastics as well as compounds in late phase clinical trials, several of which include glioblastoma-relevant drugs such as PI3K/ mTOR inhibitors, VEGFR inhibitors and met-inhibitors.

To date, we have optimized drug screening assays with the 160-compound library using the SN186 parental cell line as shown in Figure 12. Drug potency is typically assessed by determining the half maximal inhibitory concentration (IC_{50}) (*i.e.* the concentration of a drug that is required for 50% inhibition *in vitro*). With this in mind, we generated 8-point dose response curves to access the IC_{50} of each drug (Figure 12B). This same technique will be applied to SN243 and SN291 parental cell lines, and their corresponding quantized cell populations in the final year.

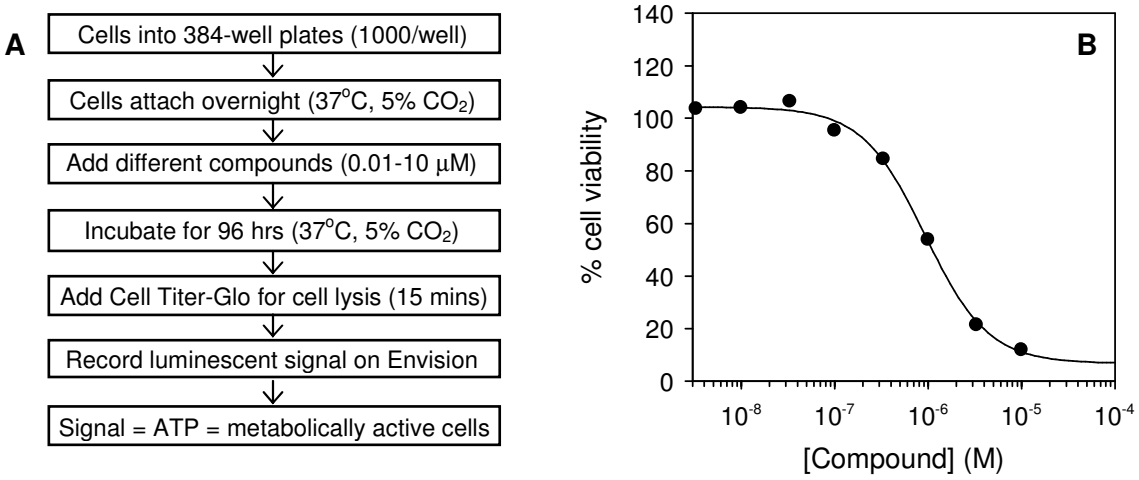


Figure 12. (A) Established protocol for high throughput chemical screens. Viable cells (metabolically active) are determined by ATP quantification, using the CellTiter-Glo luminescent cell viability assay (Promega). **(B)** Example of dose response curve used to determine IC50 value for each drug candidate.

We have encountered a significant logistical problem with the explosion of whole genome sequencing requests and capacity by our contractor Complete Genomics Inc. We have addressed these issues with them and are progressing towards the completion of this project. With large changes in staff at CGI, it has taken time to reestablish an error free system and we have been subjected to these delays. Other technical problems we encountered were solved and we do not anticipate problems with the proposed work schedule for the next 12 months. Our

efforts in whole genome sequencing have advanced since or request for a no-cost extension of the proposal and bode well for the upcoming work to be completed. This program will be completed as stated and has benefited from further significant technical developments we have made and will provide significant results over the next 12 months.

KEY RESEARCH ACCOMPLISHMENTS FOR 2013-2014

- We have established five GBM primary cell lines suitable for extensive molecular analysis (as listed in Table 1). For three of these patient samples (SN243, SN291 and SN348), we were able to get full family consent and have therefore collected blood from the patient's family members in addition to the blood and tumor specimens collected from the GBM patients (Table 2). These samples are suitable for both cellular analysis, genomic analysis and proteomic analysis.

- Quantized cell populations have been established from SN243 and SN291 parental tumor cell lines.

- Whole genome sequencing has been completed for patient SN291, to include tumor, primary cell line, five separate quantized cells and SN291 family members.

- Key mutations in the cancer genome of SN291 have been identified.

- Transcriptomic profiling of single cells from SN291 quantized populations has been performed. Distinct expression patterns of selected genes have been identified.

- Cell culture conditions for secretome analysis of quantized cells and protein extraction conditions to maximize the amount of protein for high-mass accuracy quantitative mass spectrometry have been developed.

- Deep proteome secretome analysis of individual GBM primary cells and their quantized subclones has been performed.

- Initial proteome identifications from GBM tumor samples have been made. Cancer proteome specific database strategies to identify protein mutations predicted by whole genome sequencing have been established.

- A drug screening protocol has been developed for screening SN243 and SN291 parental lines and quantized cell populations.

REPORTABLE OUTCOMES

Publication describing technical developments applied to GBM cell analysis:

Sangar V, Funk CC, Kusebauch U, Campbell DS, Moritz RL, Price ND. Quantitative proteomic analysis reveals effects of EGFR on invasion-promoting proteins secreted by glioblastoma cells. Mol Cell Proteomics. 2014 Jul 5. pii: mcp.M114.040428. PMID: 24997998.

CONCLUSION

Description of work to be performed over the next 12 months.

We will complete our whole genome sequence collection for patient and family SN243. This will complete our data collection for whole genome sequencing and provide a second extensive dataset to supplement the efforts we have made significant inroads on. We have established a data analysis pipeline wit patient SN291 and their family members and will complete this analysis.

We will complete our full transcript analysis of single quantized cells from both SN291 and SN243 to establish expressed gene and mutations identified in the whole genome sequence analysis of these sample types.

We will complete our proteomics analysis of the full sample set of expanded GBM patients and of the family members from two of these undergoing full genomic analysis. We will perform tumor proteome specific analysis of GBM primary and quantized cell lines to provide quantitative differences between these as well as detect and quantify expressed protein mutations identified from the multi-omic approach.

We will compile a candidate biomarker list of select targets derived from WGS, single cell RNA-seq and cancer proteome data of both whole cells as well as quantize cell secretomes to build targeted assays for candidate biomarker evaluation. We will construct a panel of SRM assays and analysis conditions to deploy across GBM patient and normal subject plasma samples to conduct a biomarker evaluation for early detection of GBM tumors in both a unblended and blinded analysis.

We will complete our targeted drug screening profile of GBM patient primary and quantized cells to reveal inhibitors of GBM stem cell proliferation and compile a list of drugs that have the potential to be tested in Phase-II clinical trials.

REFERENCES

(1) Hothi, P., Martins, T. J., Chen, L., Deleyrolle, L., Yoon, J. G., Reynolds, B. Foltz, G. (2012) High-throughput chemical screens identify disulfiram as an inhibitor of human glioblastoma stem cells. *Oncotarget 3*, 1124-36.

(2) Pollard, S. M., Yoshikawa K, Clarke ID, Danovi D, Stricker S, Russell R, Bayani J, Head R, Lee M, Bernstein M, Squire JA, Smith A, Dirks P. (2009) Glioma stem cell lines expanded in adherent culture have tumor-specific phenotypes and are suitable for chemical and genetic screens. *Cell Stem Cell 4*, 568-80.

(3) Clément, V., Dutoit, V., Marino, D., Dietrich, P.Y., Radovanovic, I. (2009) Limits of CD133 as a marker of glioma self-renewing cells. *Int J Cancer 125*, 244-8.

(4) Clément, V., Marino, D., Cudalbu, C., Hamou, M.F., Mlynarik, V., de Tribolet, N., Dietrich, P.Y., Gruetter, R., Hegi, M.E., Radovanovic, I. (2010) Marker-independent identification of glioma-initiating cells. *Nat Methods 7*, 224-8.

(5) Wang, J., Sakariassen, P.O., Tsinkalovsky, O., Immervoll, H., Boe, S.O., Svendsen, A., Prestegarden, L., Rosland, G., Thorsen, F., Stuhr, L., Molven, A., Bjerkvig, R., Enger, P.O. (2008) CD133 negative glioma cells form tumors in nude rats and give rise to CD133 positive cells. *Int J Cancer 122*, 761-8.

(6) Beier, D., Hau, P., Proescholdt, M., Lohmeier, A., Wischhusen, J., Oefner, P.J., Aigner, L., Brawanski, A., Bogdahn, U., Beier, C.P. (2007) CD133(+) and CD133(-) glioblastoma-derived cancer stem cells show differential growth characteristics and molecular profiles. *Cancer Res 67*, 4010-4015.

(7) Vescovi, A. L., Galli, R., Reynolds, B.A. (2006) Brain tumour stem cells. *Nat Rev Cancer 6*, 425-36.

APPENDICES

N/A