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: September 2015

TYPE OF REPORT: Final Report

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**
September 2015	Final	15June2011 - 14June2015

4. TITLE AND SUBTITLE

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

5a. CONTRACT NUMBER
W81XWH-11-1-0488

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. Charles Cobbs

E-Mail: 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
747 Broadway
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

This research proposal provides important insights into cancer mechanisms and blood biomarkers to assess progression and stratification of human glioblastoma. We have successfully established several patient-derived cell lines from glioblastoma tumors and further established a number of quantized cell populations from these parental tumor cell lines. The cell populations have been used for extensive molecular characterization, including whole genome sequencing, transcriptomic analysis, proteomics (secretome as well as N-glycocapture analysis), and high-throughput drug screening for evaluating the effectiveness of drug candidates in targeting these cell populations. Combined, this program significantly advances genomic, proteomic and single-cell technologies. Although the focus of this research is glioblastoma, the proposed tools are generally applicable to all cancer-based studies.

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**
U					USAMRMC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U	UU	**26**	

Table of Contents

INTRODUCTION

This collaborative research program delivers 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 primary brain tumor in adults which remains an incurable and rapidly fatal disease. Cancer stem cells have been implicated as the presumed cause of tumor recurrence and resistance to therapy (1-4). With this in mind, we have utilized glioblastoma patient-derived cell lines and an integrative multi-omic approach to study glioblastoma stem cell populations and their role in disease progression. This has involved the development of new strategies for advanced genome sequencing, the analyses of transcriptomes, miRNAomes and single cells as well as multiplexed quantitative protein measurements including the measurement of isoforms, and post-translational modifications. We believe this proposal has significantly advanced genomic, proteomic and single-cell technologies, and the proposed tools, which identify and quantify DNA, RNAs, proteins and cells, are generally applicable to all cancer-based studies. To accomplish these goals we pursued the following aims:

Specific Aim 1. Isolate up to 1000 cells from each of five human glioblastomas 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 glioblastoma tumors 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 glioblastoma patients with regard to their ability to stratify disease, assess disease progression and predict at an early stage glioblastoma recurrence. Eventually we will use these biomarkers to assess the effectiveness of therapy.

Specific Aim 4. Ten to 20 cells from each major quantized glioblastoma 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 to the perturbations of key glioblastoma-relevant molecules (e.g. nodal points in networks) by RNAi perturbations as well as their responses to glioblastoma-relevant drugs and natural ligands.

This is the final report for the second year of program (a second no-cost extension was awarded in Year 3 because of delays in obtaining IRB approvals and delays in whole genome sequencing). This report will summarize the work conducted over the entire research period, which has allowed us to establish tumor cell lines and define new molecular targets as markers of disease progression and patient outcome to therapy.

We believe the outcomes and deliverables of this program include: 1) a deeper understanding of human glioblastoma; 2) blood protein biomarkers for use in early diagnosis and assessment of disease progression, assessment of drug treatment effectiveness, and early detection of disease recurrence; 3) new strategies for genomic sequencing to identify relevant mutations; 4) new technologies for transcriptome, miRNAome, proteome, and single-cell analyses, and 5) the creation of quantized glioblastoma cell lines that can be used for general molecular characterization and to evaluate the effectiveness of existing drugs in reacting with these cell types.

BODY

Specific Aims 1, 2 and 4.

Quantized glioblastoma cell populations. The Ivy Center for Advanced Brain Tumor Treatment at the Swedish Neuroscience Institute collected tumor tissue eligible for this program from over fifty glioblastoma patients over the entire research period. We used a well-established protocol to generate multiple primary tumor cell lines from the tissue specimens. Importantly, our patient-derived tumor cell lines preserve the stem cell phenotype; namely self-renewal, the ability to differentiate into different cell types, and the ability to generate tumors *in vivo* (Figure 1). Tumor heterogeneity and individual patient responses are principal contributing factors to the difficulty in designing general treatment regimens for glioblastoma patients. The inherent heterogeneity of glioblastoma is reflected in tumor stem cells, which differ in their proliferative potential, tumor-initiating ability and therapeutic responses, and more closely resemble the parent tumor both genotypically and phenotypically (5). With this in mind, we believe the evaluation of glioblastoma-derived stem cell populations using an integrative multi-omic approach (i.e. genome sequencing, the analyses of transcriptomes, miRNAomes and single cells, as well as multiplexed quantitative protein measurements) is essential to understanding glioblastoma disease progression.

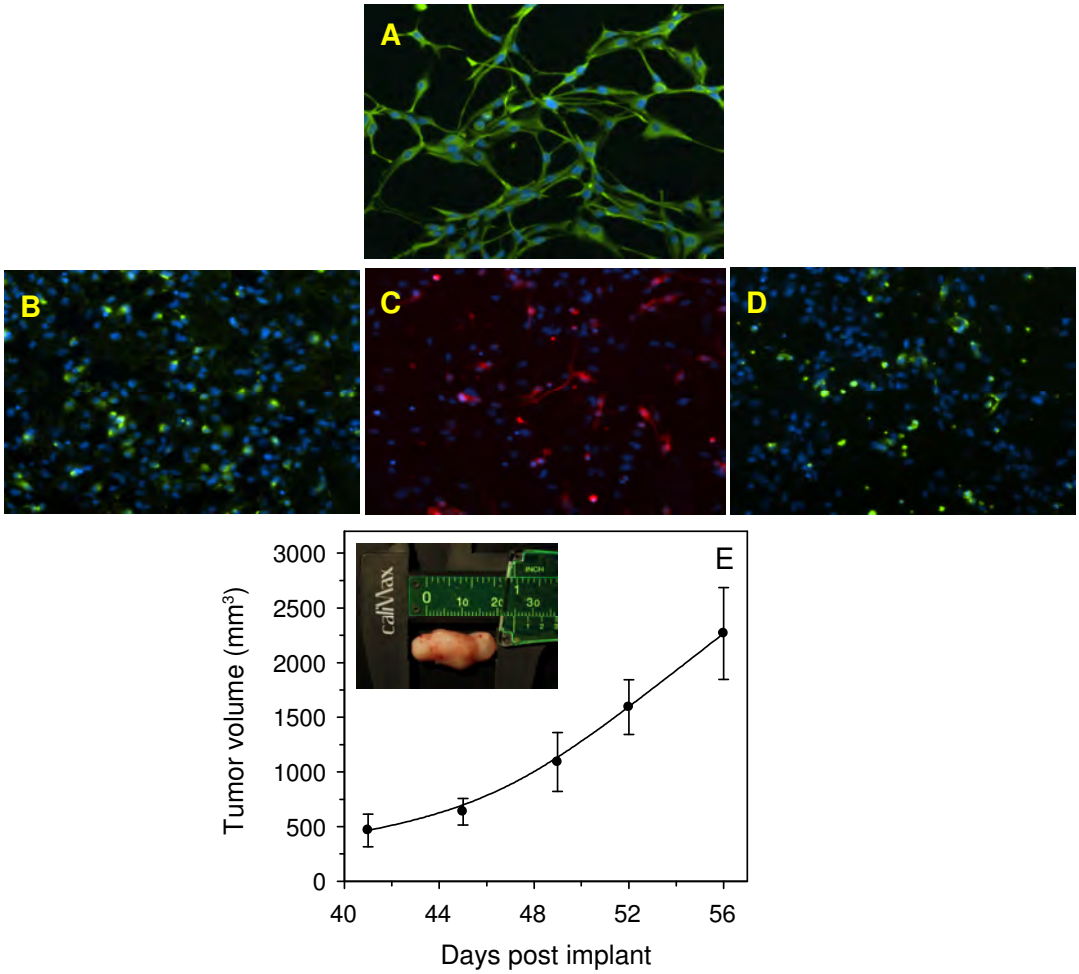


Figure 1. (A) Adherent cultures of glioblastoma-derived tumor cells. Tumor tissue is dissociated immediately after surgical resection, and single cell suspensions are plated in serum-free NeuroCult® NS-A media with B27, epidermal growth factor (EGF) and fibroblast growth factor (FGF-2). Cultures established using this method fulfill the accepted criteria for cancer stem cells, namely self-renewal, multipotency, and tumor-initiating ability *in vivo*. Immunostaining for differentiation markers: **(B)** GFAP/astrocytes, **(C)**, TUJ-1/neurons and **(D)** O4/oligodendrocytes. Cells were grown in NS-A media without growth factors (EGF and FGF-2) for 2 weeks. Primary antibodies were from R&D Systems and goat secondary antibodies conjugated to Alexa dyes were from Invitrogen. DAPI (Sigma) was used as the nuclear counterstain. Images were acquired using a Nikon Ti-U inverted fluorescence microscope liked to a DS-U2 camera. **(E)** *Main panel:* Tumor formation *in vivo* confirms the presence of stem cell populations within the heterogeneous cell culture as thus suitable of cell lines for the proposed research. *Inset:* tumor mass removed from xenografts injected with patient-derived cultures.

We transferred several of our glioblastoma stem cell 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.

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 glioblastoma patients used in this study.

A number of quantized cell populations were successfully established from the corresponding parental cell lines (Figure 2). 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, clonal cultures which exhibited distinct morphological phenotypes were established. Given that each clone presumably carries a uniform genome, it is suitable for whole genome sequencing. These cell populations thus serve as the foundation for genomic, transcriptomic, and proteomic studies.

In particular, the glioblastoma specimens SN243 and SN291, for which we had consenting family members, were selected for complete molecular analyses. We collected blood (processed as plasma and peripheral blood mononuclear cells [PBMCs]; Figure 3), from both SN243 and SN291 patients, and their family members (Table 2). This completed the specimen cohort required for molecular analyses at the genome, transcriptome, miRNAome and proteome levels (Specific Aims 1, 2 and 4). Five clones were selected from each patient for subsequent 'omic analysis.

Whole transcriptomics analysis was performed on selected clones from both patient samples in order to evaluate molecular heterogeneity at the transcript level. The observed cell population distribution pattern was consistent with the single cell gene expression analysis. From these combined analyses, a panel of genes that potentially function as glioblastoma subpopulation-specific markers was established, for further evaluation in SRM-based targeted proteomics assays (see PI Moritz report).

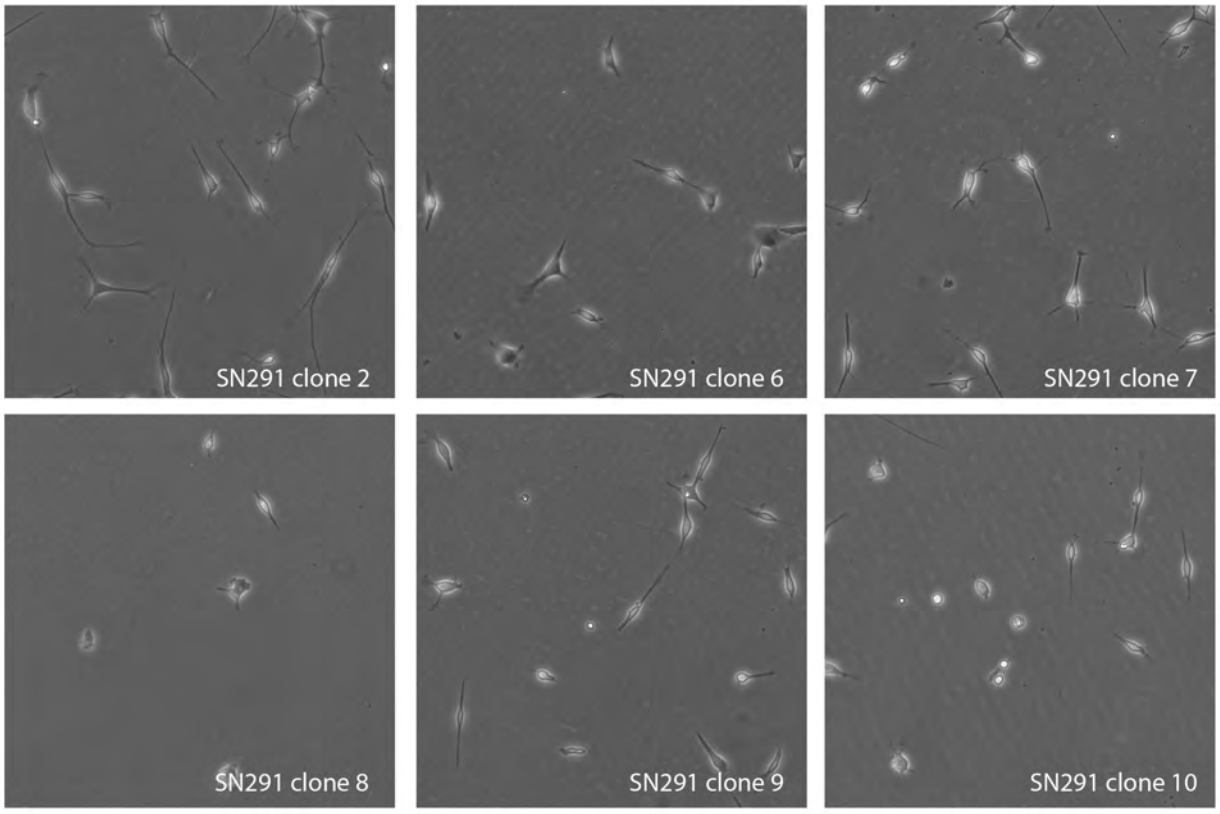


Figure 2. Establishment of single cell clonal cultures from glioblastoma patient SN291. A total of 12 clonal cultures were generated from this patient sample. In brief, cells were cultured on plates coated with laminin and grown under serum-free conditions with stem cell media supplement with B27, N2 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.

Whole genome sequencing. As proposed, two patient families (SN243 and SN291) were selected for whole genome sequencing analyses (Figure 4). DNA samples from tumor tissue, parental cell line, five subclones, and the genomes of family members were prepared for whole genome sequencing at Complete Genomics. High-quality whole genome sequences were obtained to include the patient (from PBMCs), family members, as well as the original tumor tissue, the parental tumor cell line, and five isolated single cell subclones (Figure 5). The overall goal of the whole genome sequencing is to provide insight into the mutational landscape of individual clones derived from the tumors with relation to the heterogeneous whole tumor genome and correction with the genomes of parents and offspring. Analysis of the whole genome sequencing data was completed using the family genomics pipeline at ISB (see PI Moritz report).

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.

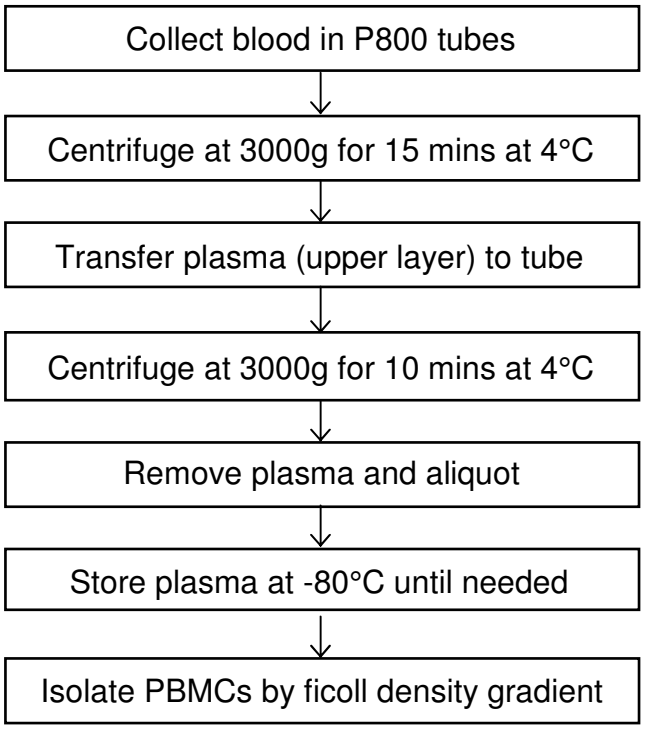


Figure 3. Established protocol for the isolation of human plasma and PBMCs from whole blood samples.

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, based on comparison of the genome coverage signal to a pre-computed "reference coverage profile", was developed (Figure 6). For analyzing the genomes, a reference coverage profile based on 106 normal genomes (all obtained from blood samples and excluding the currently studied genomes) was generated. The genomes were normalized to this reference profile and used to identify regions of coverage that were lower or higher than expected (see PI Moritz quarterly report).

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.

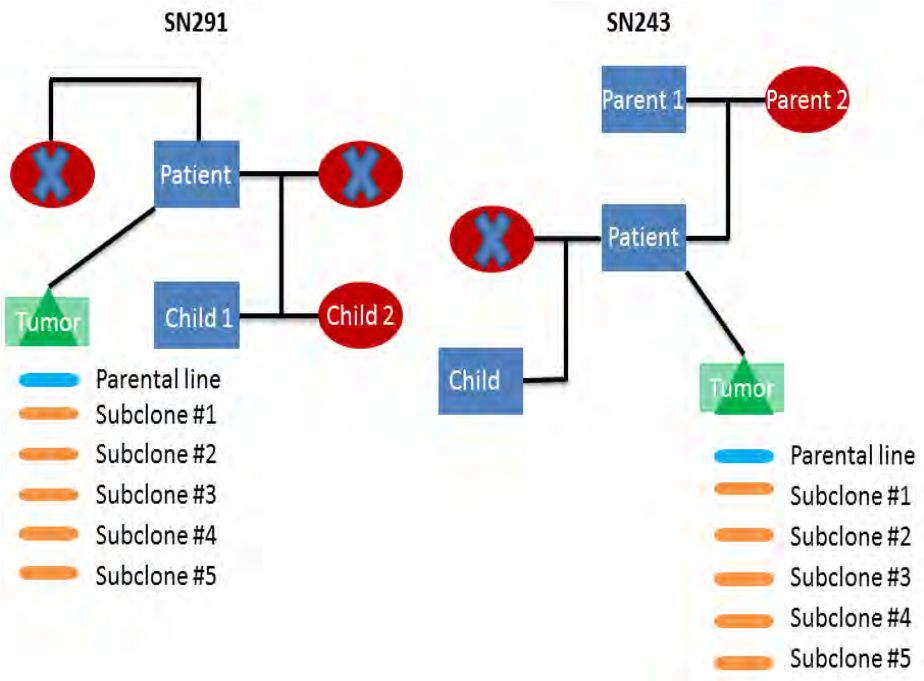


Figure 4. Patient selection and family history of samples used for whole genome sequencing.

.

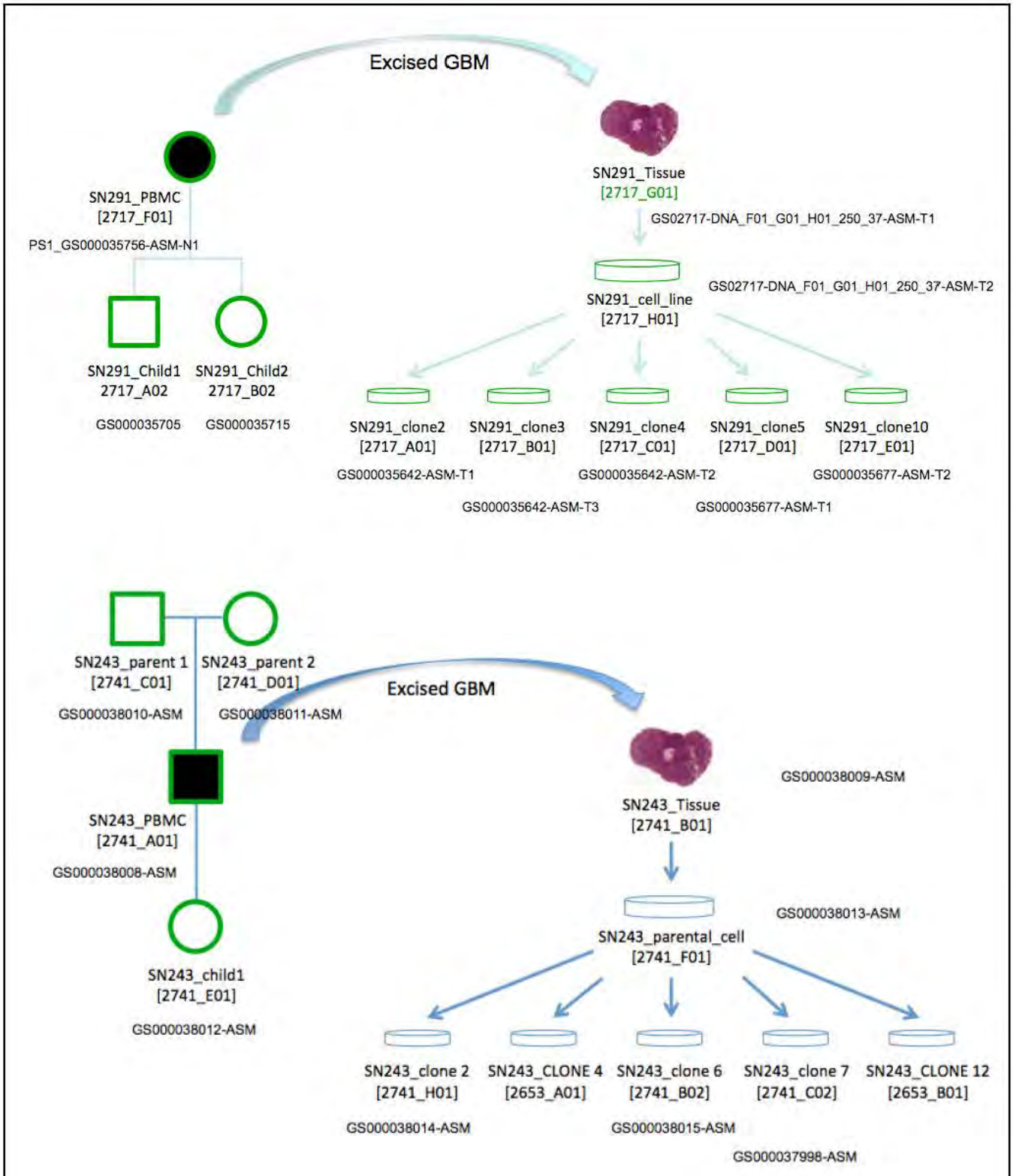


Figure 5. Genome dataset for patients SN243 and SN291. The descriptive identifier, the vendor sample identifier (square brackets), and the vendor assembly identifier are shown for each sample.

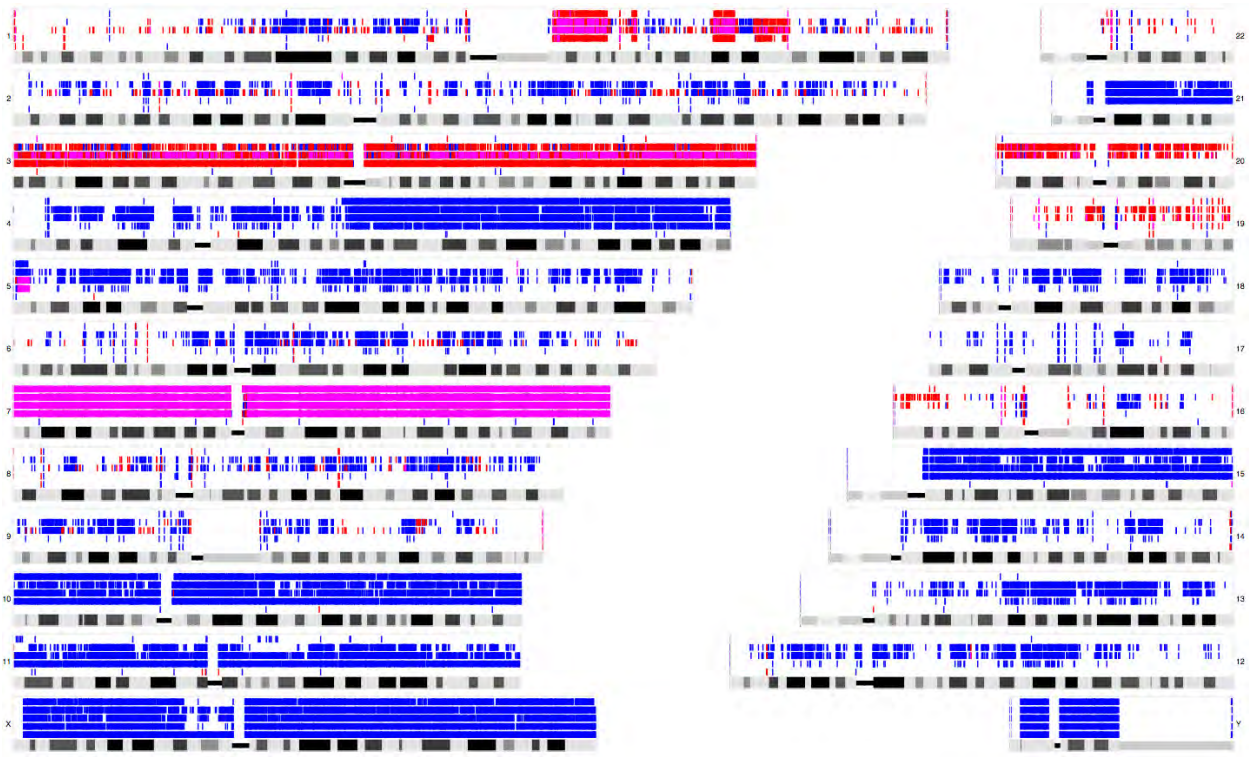


Figure 6. Computed karyotype. For each chromosome, the computed copy numbers observed for SN243's PBMC genome, cell line and three subclones are shown (bottom to top). Blue denotes deletions (haploid), red represents expansions (triploid, with magenta representing tetraploid or higher). The sex chromosomes are haploid since the subject is a male.

Variant analysis. 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 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%. Ingenuity's knowledgebase was used to select cancer driver variants directly affecting genes known to be involved in glioblastoma. A number of interesting gene mutation candidates were identified (Figure 8). Of particular interest is a stop-gain SNV in the RAD51B gene, present in heterozygous form in the genome of the SN291 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.

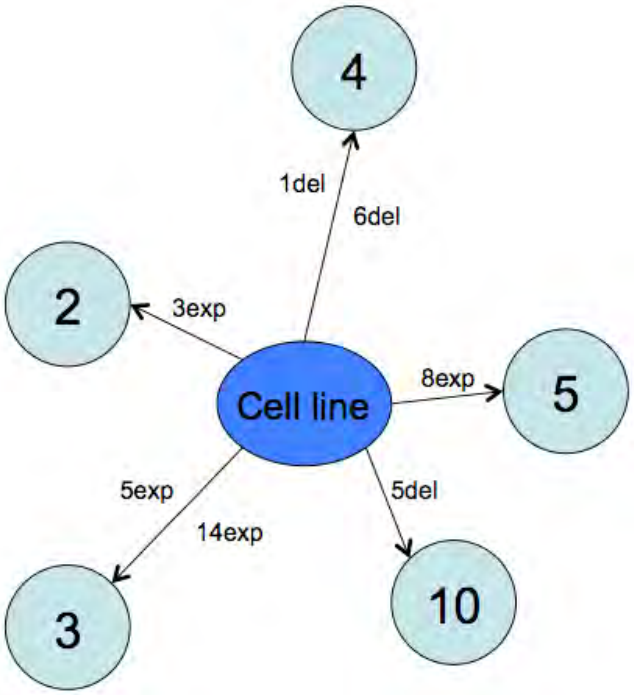
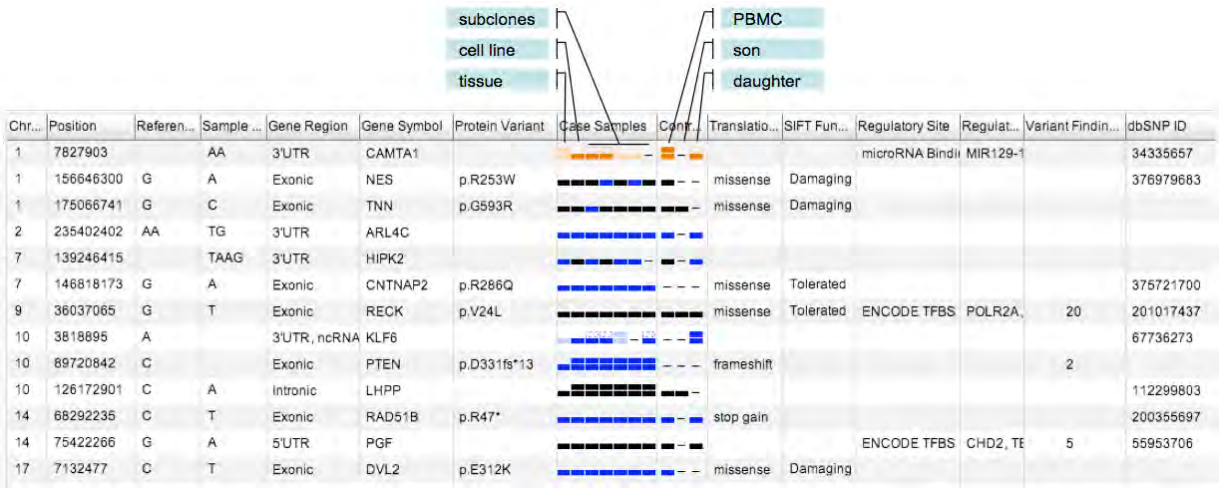


Figure 7. Relationships between the SN291 subclones and the parental cell line. Each subclone presents a very small number of private aneuploidies (deletions or expansions).

Transcriptomic analysis of glioblastoma subclone heterogeneity through RNA-Seq. One of the aims of this project is to identify candidate blood biomarkers in the bloods of glioblastoma patients (Specific Aim 3), based on transcriptomic and shotgun proteomic analysis of the quantized cell populations derived from SN243 and SN291 tumor tissues. For this purpose, high quality total RNAs were extracted from the parental cell lines and a total of 13 tumor clones (six for SN291 and seven for SN243). Between 16 and 26 million pairs of 51er nucleic acid reads were produced on the Illumina HiSeq 2000 instrument (Table 3). Our collaborators at ISB have analyzed the RNA-seq datasets utilizing data analysis programs such as Top Hat and Cuff links, with over 95% of them being mapped to the human genome (see PI Moritz report).

Principle component analysis and network mapping. Principle component analysis was performed on the single cell transcriptomes. As shown in Figure 9, several distinct cell clusters were identified for SN291. Our collaborators at ISB used their previously published work (6) 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.



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 variants. Orange and blue denote gain and loss of function, respectively.

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

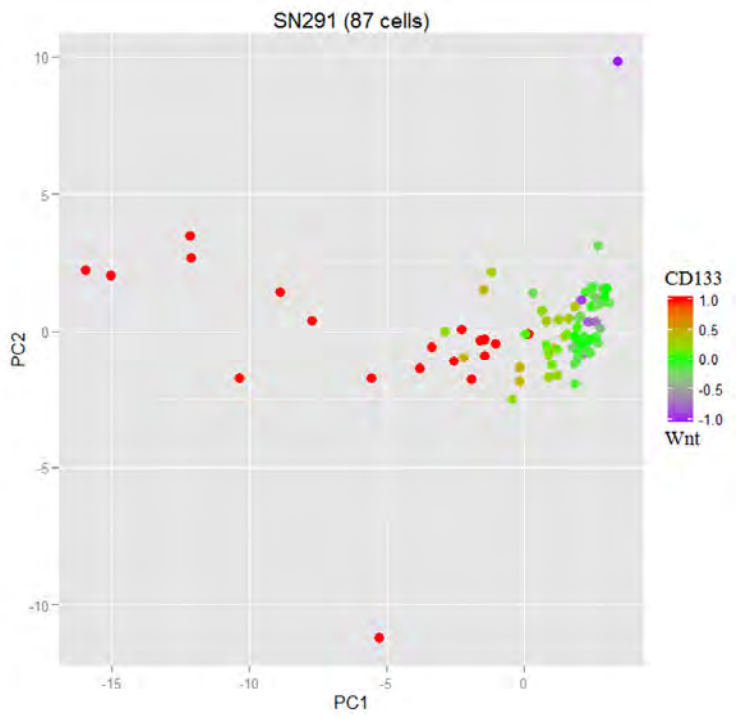


Figure 9. Principle component analysis of RNA-seq data generated from individual cells from the SN291 parental culture. Each dot represents a single cell. Color gradient indicates enrichment score for either published CD133+ gene signature (6) or Wnt pathway genes.

Specific Aim 3.

Proteomic analysis of quantized cell populations form established glioblastoma tumor cells. Parental and quantized cell populations were expanded for proteomic analysis and protocols for the stringent analysis of these samples were developed (incorporating genomic information obtained in whole genome sequencing for the establishment of candidate protein biomarkers). New growth conditions had to be established for these cells to allow for the elimination of extraneous protein from additional cell growth components and fetal bovine serum (FBS) present in the culture medium. Elimination of extraneous proteins was necessary for the identification of proteins secreted directly from the quantized cell populations. For the analysis of secreted proteins, cells were therefore grown in FBS free medium for 24 hours prior to collection in unsupplemented medium.

Proteomic data collection of all samples for glioblastoma biomarker target selection has been completed. This includes the secretome analysis as well as the N-glycocapture analysis of established glioblastoma quantized parental cell lines and subclones with methods developed at ISB (Table 4 and Table 5).

The generated 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 the 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, our collaborators at ISB generated an extended cancer genome specific database 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.

Biomarker candidates derived from this discovery proteomic analysis were correlated with the data derived from the transcriptome and whole genome analysis to define the candidates for targeted quantitative proteomic selected-reaction monitoring (SRM) analysis.

To perform SRM analysis, SRM assays for each protein target (and possible variants) are extracted from the ISB unique Human SRMAtlas website, a compendium of over 170,000 SRM assays that covers >99.9% of the Human proteome (www.srmatlas.org). To perform SRM analysis of glioblastoma differentially abundant proteins, plasma samples from normal and glioblastoma patients are first depleted from the top 14 most abundant human plasma proteins by immunoaffinity chromatography. Samples are then digested with trypsin and analyzed on an Agilent triple quadrupole mass spectrometer. Data is analyzed using Skyline software to measure the abundance of proteins by their SRM assay, an assay that measures protein signatures by proteotypic peptide quantitatively. SRM assays are measured as a multiplexed analysis allowing up to 200 peptides to be measured in a single analysis (Figure 10). Proteomic analysis performed in this manner allows the identification of differentially abundant protein candidates for SRM assay selection.

The overall aim of this effort is to evaluate glioblastoma specific tumor markers in a larger pool of blood plasma samples. To allow the completion of this aim, the Ivy Center collected and processed plasma from 100 glioblastoma patients. The plasma specimens were then transferred to ISB to assess candidate blood biomarkers useful in early diagnosis, stratification, and assessment of glioblastoma progression, and early detection of disease recurrence (see PI Moritz report).

Database construction for cancer derived mutational proteome analysis. The standard method for identifying proteins in a mass spectrometry experiment involves the use of a whole proteome database to compare sequence information, which is processed *in silico* and compared to the experimentally observed spectra. The database is meant to represent every possible polypeptide sequence fragment from the subject organism, to afford the best chance of correctly interpreting each experimental spectrum. The quantized cell glioblastoma project has identified numerous mutations from whole genome sequencing, many of which would encode novel polypeptide sequences that would not be identifiable using traditional proteomics approaches. Since it is clearly infeasible to consider every possible rearrangement of the human genome and resultant modified peptides, our collaborators at ISB devised a method to encode these variable sequences in a modified whole proteome database in a manner that will enable the detection of the fragmentation spectra from such modified peptides.

Essentially any novel polypeptide sequence resulting from observed genetic rearrangements are appended to the canonical sequence for that particular gene product, with a reasonable amount of flanking sequence as context, to account for missed enzymatic cleavages. Each ‘cassette’ consisting of a modified sequence plus context is separated from each other and the original sequence by an asterisk character. The asterisk is treated as a hard-stop boundary by most search engines, as well as the TPP software used to interpret the results in a statistically valid manner, so there is no chance of introducing spurious mutations. This allows us to encode virtually all likely sequences in a relatively compact and non-redundant manner, both desirable qualities to keep the search times reasonable and limiting the protein inference problem. This database has been constructed using the whole genome sequencing data for SN243 and SN291.

Proteomics - Identified Peptides and Proteins in Secretome and Glycocapture Preparations						
Secretome Analysis using Variant Database						
iProphet	**SN291_P**	**SN291_S2**	**SN291_S3**	**SN291_S4**	**SN291_S5**	**SN291_S10**
min probability	0.9	0.9	0.9	0.9	0.9	0.9
error	0.006	0.006	0.007	0.007	0.007	0.006
spectra	13627	14862	9505	11585	11136	15244
unique peptides	8368	9000	5899	7561	6827	9124
unique stripped peptides	8314	8957	5858	7517	6827	9082
proteins	1660	1916	1294	1617	1534	1931
single hits	456	501	397	497	463	534
ProteinProphet						
min probability	0.9	0.9	0.9	0.9	0.9	0.9
error	0.006	0.006	0.006	0.006	0.007	0.006
protein (group) entries	1194	1406	920	1169	1115	1428
single hits	130	208	96	138	149	221
Glycocapture Analysis using Variant Database						
iProphet	**SN291_P**	**SN291_S2**	**SN291_S3**	**SN291_S4**	**SN291_S5**	**SN291_S10**
min probability	0.9	0.9	0.9	0.9	0.9	0.9
error	0.006	0.008	0.007	0.008	0.008	0.009
spectra	7692	5791	8686	10396	6232	2682
unique peptides	1908	1488	2124	2603	1702	987
unique stripped peptides	1421	1148	1654	2084	1289	810
proteins	574	525	645	801	551	420
single hits	70	88	86	112	76	96
ProteinProphet						
min probability	0.9	0.9	0.9	0.9	0.9	0.95
error	0.008	0.009	0.008	0.008	0.01	0.007
protein (group) entries	503	439	509	659	462	337
single hits	198	191	196	241	189	151

Table 4. Proteome analysis of SN291 parental (P) cell line and subclones (S) of the secretome and glycoproteome.

Secreted		Parental cell line	Clone 2	Clone 4	Clone 3	Clone 5	Clone 10	Child 1	Child 2
Sample	Total	GS02717-DNA_F01_G0	GS000035642-.	GS000035642-	GS000035642-.	GS000035677-.	GS000035677-	GS000035705-.	GS000035715-/
Sn291P	48	46	44	44	45	44	45	37	45
Sn291S2	44	44	44	44	44	43	43	37	40
Sn291S3	24	24	24	23	24	23	24	19	21
Sn291S4	34	34	34	33	34	33	34	28	29
Sn291S5	34	32	32	31	32	31	32	24	29
Sn291S10	35	35	34	34	34	34	34	25	30
Glycocapture		Parental cell line	Clone 2	Clone 4	Clone 3	Clone 5	Clone 10	Child 1	Child 2
Sample	Total	GS02717-DNA_F01_G0	GS000035642-.	GS000035642-	GS000035642-.	GS000035677-.	GS000035677-	GS000035705-.	GS000035715-/
Sn291P	12	12	11	11	11	11	11	9	8
Sn291S2	9	9	9	9	9	9	9	7	7
Sn291S3	15	14	13	13	13	13	14	9	11
Sn291S4	16	15	13	13	13	13	13	11	12
Sn291S5	11	9	9	9	9	9	10	6	7
Sn291S10	8	8	8	8	8	8	8	6	6

Table 5. Variant sequence detection of glioblastoma parental cell lines and subclones.

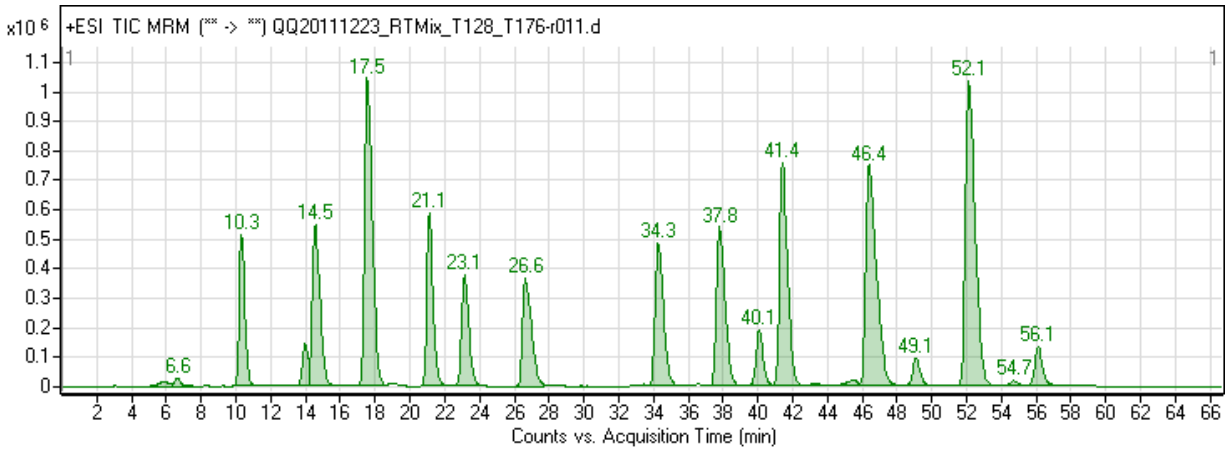


Figure 10. SRM proteomic analysis of proteotypic peptides.

Specific Aim 5.

Responses of quantized cell populations to glioblastoma-relevant drugs. We have completed high-throughput drug screening using a 160 compound library (as summarized in Figure 11). This drug 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 (Table 6). 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 against SN243 and SN291 parental cell lines, as well as SN243 subclones (SN243-2, SN243-4, SN243-6, SN243-7 and SN243-12). IC_{50} values were determined by fitting data to the standard four-parameter sigmoidal dose response curve. The IC_{50} values were then used to identify potential drug candidates that have potency against the glioblastoma cell populations tested.

66.2% (106/160) compounds did not inhibit glioblastoma cell proliferation (Figure 12A). 19.4% (31/160) compounds had IC_{50} values in the high micromolar range (~8 – 63 μM) (Figure 12B). 14.4% (23/160) have IC_{50} values in the nanomolar to lower micromolar range

Activity/ Function	Total
Antineoplastic	27
Multi-kinase inhibitor	7
mTOR / PI3K inhibitor	22
Protein kinase C	6
Bcl-2 inhibitor	3
CDK inhibitor	8
MEK1/2 inhibitor	6
VEGFR inhibitor	8
EGFR inhibitor	3
PARP inhibitor	2
Interleukin inhibitor	1
c-Met inhibitor	4
Statin	1
NF-kB inhibitor	2
Src/ Abl inhibitor	3
CHK inhibitor	4
HDAC inhibitor	5
Survivin inhibitor	2
Anti-inflammatory	2
Proteasome inhibitor	2
Hedgehog (Hh) inhibitor	4
JAK inhibitor	2
IGF-1R inhibitor	2
HER1/EGFR tyrosine kinase inhibitor	2
ALK inhibitor	2
AKT inhibitor	1
Heat Shock Protein 90 inhibitor	2
B-raf enzyme inhibitor	2
FLT3 tyrosine kinase inhibitor	3
Antimetabolite	2
Polo-like kinase 1 inhibitor	2
Retinoic acid receptor	3
Farnesyltransferase inhibitor	2
Other	13
	160

Table 6. Oncology library.

(~0.03 – 7 μM). In particular, we found that some drugs had a differential response on the parental cell lines versus the subclone populations (Figure 12C). The lead drug candidates (*i.e.* those with the highest degree of potency) are listed in Table 7, and should be evaluated further for potential use in the treatment of glioblastoma patients.

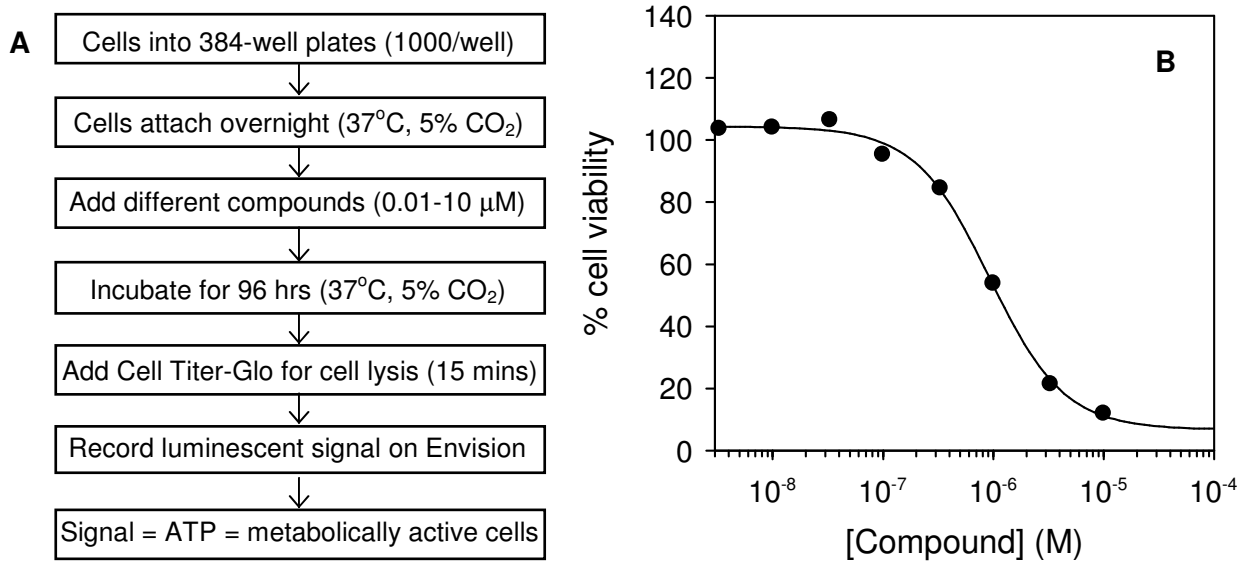
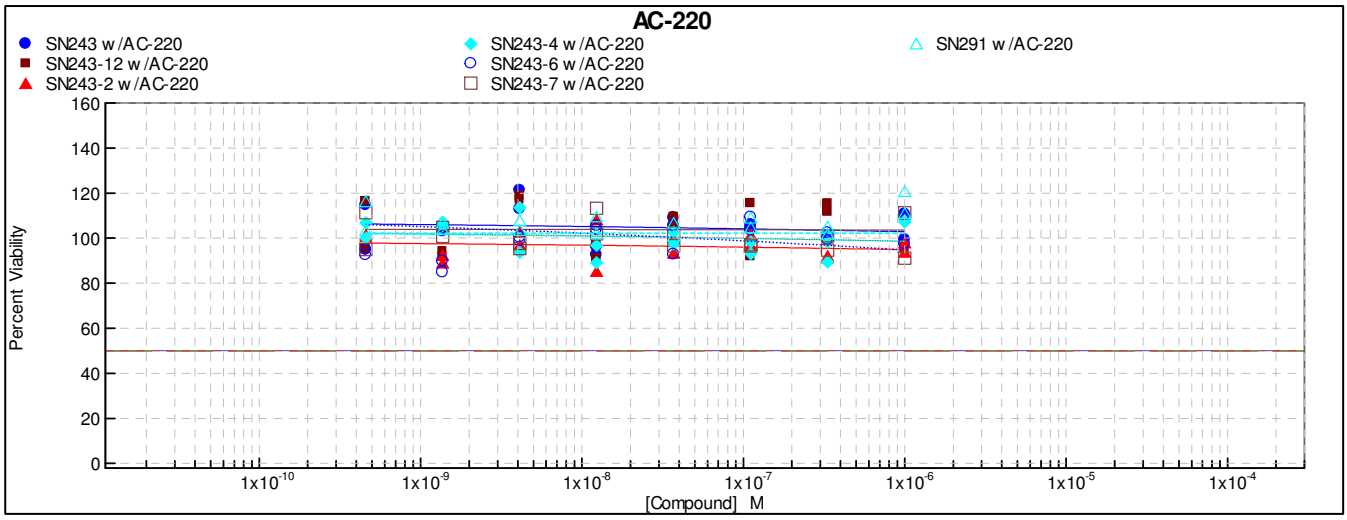
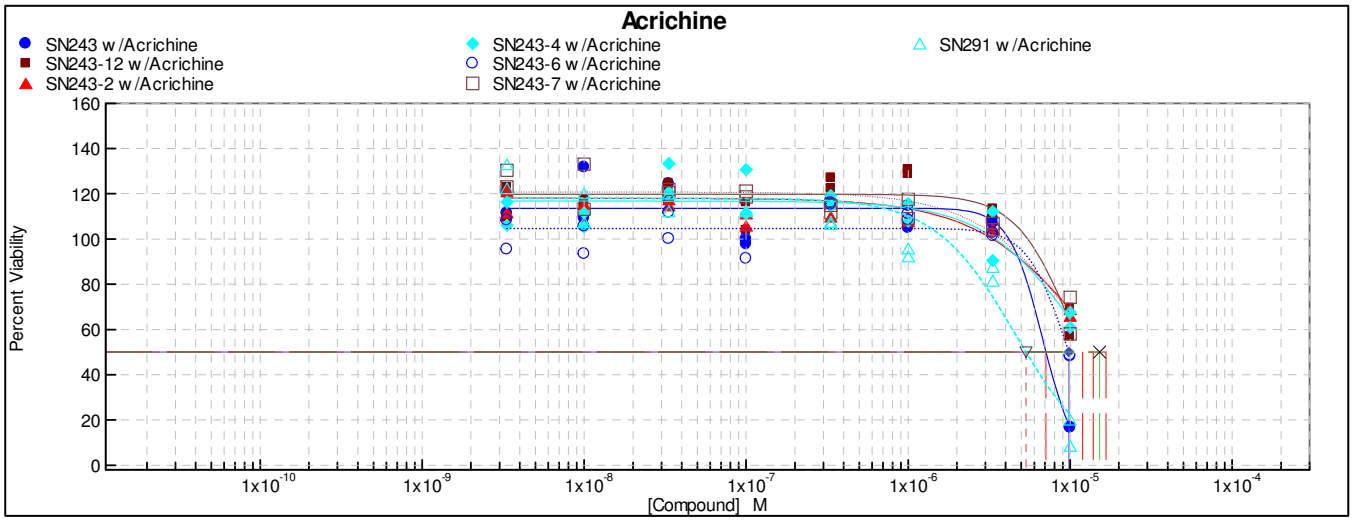


Figure 11. (A) Established protocol for high throughput chemical screens (7). Viable cells (metabolically active) are determined by ATP quantification, using the CellTiter-Glo luminescent cell viability assay (Promega). **(B)** Example of dose response curve used to determine IC_{50} value for each drug tested against the glioblastoma cell populations.

A



B



C

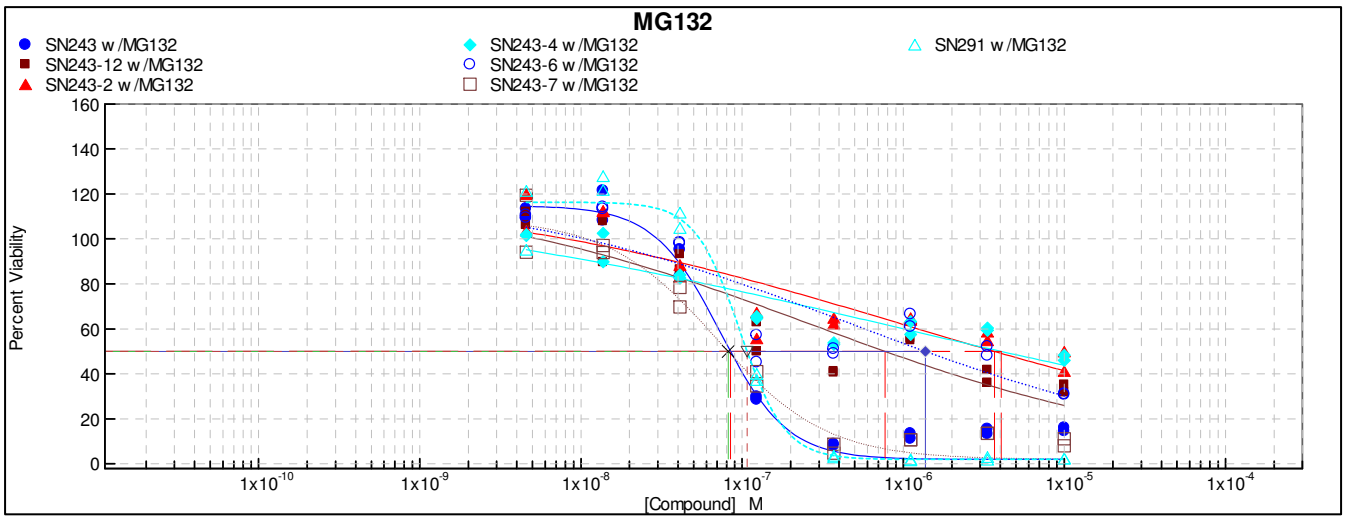


Figure 12. Dose response curves for SN291, SN243 and SN243 subclones.

Compound	Function / Activity	Current Status
Fenretinide	Synthetic retinoid deriverative: accumulation of reactive oxygen species (ROS) promotes apoptosis	38 clinical trials: Phase II/III for solid tumors, head & neck, acute myeloid leukemia (AML)
Obatoclax	BCl-2 inhibitor: induces apoptosis in tumor cells, experimental drug for various cancers	18 clinical trials: Phase I/II for leukemia, lymphoma, lung cancer
YM-155	Survivin inhibitor: survivin protein (BIRC5) highly expressed in human tumors, prevents apoptosis	11 clinical trials: Phase I/II for breast, melanoma, lymphoma, prostate, solid tumors
TG-101348 (SAR302503)	JAK2 (Janus kinase 2) inhibitor: blocks JAK-STAT signaling leading to induction of apoptosis	Developed for myeloproliferative diseases. 11 clinical trials: 2 for solid tumors, others for myelofibrosis
AP24534 (Ponatinib)	Multi-Kinase inhibitor: targets Abl, PDGFRa, VEGFR2, FGFR1, Src	FDA approved for CML & ALL (temporarily suspended / partial hold on new trials due to side effects). 16 clinical trials: solid tumors, head & neck, thyroid.
Pp-242 (TORKinib)	selectivity as mTOR inhibitor over other PI3K kinases, augments TRAIL-induced apoptosis of cancer cells	Phase I clinical trials
ARQ-197 (Tivantinib)	c-Met receptor tyrosine kinase inhibitor	41 clinical trials: Phase II for various solid tumors including head & neck
PKC-412 (Midostaurin)	Multi-Kinase inhibitor: potential antiangiogenic and antineoplastic activity	20 clinical trials: Phase II/III for AML, myelodysplastic syndromes (MDS), rectal cancer
Tanespimycin (17-AAG)	Heat shock protein 90 (HSP90) inhibitor: HSP90 is a chaperone protein implicated in oncogenesis	53 clinical trials: Phase I/II for various solid tumors & hematologic malignancies
NVP-AUY-922	Heat shock protein 90 (HSP90) inhibitor: chaperone protein implicated in oncogenesis	26 clinical trials: Phase I/II for various solid tumors & hematologic malignancies
BMS-754807	Insulin like growth factor 1 receptor (IGF-1R) inhibitor	6 clinical trials: Phase I/II for metastatic solid tumors, breast cancer
PIK-75	PI3K inhibitor: moderately selective for p110α isoform compared to p110β, p110δ and p110γ	Many similar PI3K inhibitors are in clinical trials
Staurosporine (AM-2282)	Protein kinase C inhibitor, induces apoptosis	33 clinical trials: Phase I/II for solid tumors and AML

Table 7. Lead drug candidates identified from the 160 compound library.

KEY RESEARCH ACCOMPLISHMENTS

In summary, the following have been established from this research program:

- **Tumor collection:** The Ivy Center for Advanced Brain Tumor Treatment at the Swedish Neuroscience Institute collected eligible tumor tissue from over fifty glioblastoma patients.

- **Primary glioblastoma cell lines:** tissue processing techniques were refined to allow for the routine establishment of glioblastoma patient-derived primary cell lines suitable for the isolation of quantized cells.

- **Quantized glioblastoma cell populations:** single cell gene expression assays identified quantized cell populations in parental glioblastoma cells. Several quantized cell populations have been established for molecular studies.

- **Family sequencing:** we were able to consent three families for whole genome sequencing and blood plasma collection for the downstream proteomic analysis of defined glioblastoma targets. Two of the three (SN243 and SN291) were selected for the study as previously described.

- **Proteomic studies:** Developed 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.

- **Methodologies for whole genome sequencing, transcriptomics, and proteomics analyses:** have been applied to quantized cell populations from two patient samples to yield promising data. A panel of genes that potentially function as glioblastoma subpopulation-specific markers has been established for SRM-based targeted proteomics. Cancer proteome specific database strategies to identify protein mutations predicted by whole genome sequencing have been developed (see PI Moritz report).

- **Improvement of proteogenomics workflow:** extended analysis to appropriately interpret multi-nucleotide variants. Developed software code to properly account for heterozygous non-reference alleles. Extended the expected variant peptides to include neighboring sequences (~30 aa before and after the variant peptide) to enhance the ability to detect variant peptides in the presence of incomplete tryptic digestion. Analyzed a large set of genomes (>7300 whole genomes) to derive statistics on how frequently the variant peptides are observed in the population.

- **Blood biomarker studies:** The Ivy Center for Advanced Brain Tumor Treatment collected blood samples from 100 glioblastoma patients to assess candidate blood biomarkers in a large patient cohort using SRM assays at ISB.
- **Glioblastoma relevant drugs:** we optimized high-throughput screening methodology to profile drug responses of the quantized cell populations, and identified several anti-glioblastoma agents.

REPORTABLE OUTCOMES

We have reported our work originating from the efforts described here in a publication describing some of our technical developments applied to glioblastoma 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

Through our collaboration with ISB, we have successfully completed whole genome sequencing of quantized glioblastoma populations, patient tumor, PBMCs from patient, and PBMCs from each family member selected from SN291 and SN243. Subsequent analyses of the whole genome sequencing data and transcript data was possible through the in-house expertise available at ISB. Using the curated list of variants in each genome, we were able to produce final versions of genome- and cell-specific proteome databases against which to analyze proteome data (Specific Aims 1, 2 and 4).

Transcript analysis of single quantized cells from SN243 and SN291 allowed the generation of a ranked list of differentially expressed proteins from both cell surface and expected cell membrane proteins. A ranked list of transcripts derived from this analysis was combined with the proteomic data on SN243 and SN291 secretome and N-glycocapture to derive a final list of ranked proteins for SRM analysis. These were used to define tumor proteome specific targets for glioblastoma biomarkers in blood, which included quantitative differences in proteins determined and supplemented with quantified protein mutations identified from the multi-omic approach. Further validation of biomarker candidates was performed in a large glioblastoma patient cohort (Specific Aim 3).

In addition, we have identified several glioblastoma-relevant drugs with potency against glioblastoma parental lines and quantized cell types (Specific Aim 5).

We believe this program has significantly advanced genomic, proteomic and single-cell technologies, as originally proposed, and enabled the commencement of hypothesis-driven integrative systems approaches to cancer.

REFERENCES

(1) Bao, S., Wu Q, McLendon RE, Hao Y, Shi Q, Hjelmeland AB, Dewhirst MW, Bigner DD, Rich JN. (2006) Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature 444*, 756-760.

(2) Lu, C., Shervington A. (2008) Chemoresistance in gliomas. *Mol Cell Biochem 312*, 71-80.

(3) Stupp, R., Mason, W.P., van den Bent, M.J., Weller, M., Fisher, B., Taphoorn, M.J., Belanger, K., Brandes, A.A., Marosi, C., Bogdahn, U., Curschmann, J., Janzer, R.C., Ludwin, S.K., Gorlia, T., Allgeier, A., Lacombe, D., Cairncross, J.G., Eisenhauer, E., Mirimanoff, R.O; European Organisation for Research and Treatment of Cancer Brain Tumor and Radiotherapy Groups; National Cancer Institute of Canada Clinical Trials Group. (2005) Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med 352*, 987-996.

(4) Brandes, A. A., Fiorentino MV. (1996) The role of chemotherapy in recurrent malignant gliomas: an overview. *Cancer Invest 14*, 551-9.

(5) Lee J, Kotliarova S, Kotliarov Y, Li A, Su Q, Donin NM, Pastorino S, Purow BW, Christopher N, Zhang W, Park JK, Fine HA. (2006) Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer Cell* 5, 391-403.

(6) Yan X, Ma L, Yoon JG, Diercks A, Foltz G, Price ND, Hood LE, Tian Q. (2011) A CD133-related gene expression signature identifies an aggressive glioblastoma subtype with excessive mutations. *PNAS* 108(4), 1591-6.

(7) 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.