Characterizing Intratumoral Heterogeneity: A Role for Cell Plasticity

Scholarly Report submitted in partial fulfillment of the MD Degree at Harvard Medical School

Date: 01 March 2016

Student Name: Teodora Kolarova, B.S.

Scholarly Report Title: Characterizing intratumoral heterogeneity: A role for cell plasticity

Mentor Name(s) and Affiliations: Joan Brugge, PhD, Department of Cell Biology, Harvard Medical School

Collaborators, with Affiliations: Suha Naffar-Abu Amara[1], Elizabeth Patricia Lunsford[1], Carina Hage[1], Laura Selfors[1], Kripa Ganesh[1], Marco Leung[2], Nick Navin[2], Gordon Mills[2]

[1]Department of Cell Biology, Harvard Medical School

[2]MD Anderson Cancer Center, Houston, TX

ABSTRACT

Title: Characterizing intratumoral heterogeneity: A role for cell plasticity

Purpose: While most human tumors are thought to be derived from a single cell, tumor cell populations demonstrate considerable heterogeneity in phenotype and behavior. During characterization of single-cell clonal populations of a primary clear cell ovarian carcinoma, we identified a particularly tumorigenic clone (Clone 31) with epithelial and mesenchymal traits suggesting a possible multi-potent progenitor. We thus wanted to determine whether tumor initiating activity and other cell properties were restricted to progenitor, epithelial-like or mesenchymal-like cells.

Methods: We generated single-cell subclones of Clone 31 using FACS. These subclones were subsequently characterized phenotypically, functionally and genetically. *In vitro* studies included monolayer proliferation, soft agar colony formation and E-cadherin and vimentin expression. *In vivo* tumor initiating ability was tested using an immunocompromised mouse model. Lastly, each subclone underwent copy number analysis, deep exome sequencing of a panel of genes and RNA sequencing to elucidate mechanism for functional diversity.

Results: While Clone 31 displayed a strong ability to initiate tumor growth *in vivo,* this property was not universally retained among the subclones. While our study was designed to identify processes and properties that correlated with tumorigenicity, we were unable to distinguish any that clearly correlated with this activity. However it is noteworthy that only subclones with mixed epithelial and mesenchymal morphology formed solid tumors *in vivo*, which Clone 31 is incapable of doing. Further, subclone expression data is suggestive of epithelial mesenchymal plasticity among tumorigenic subclones and thus warrants further investigation.

Conclusions: The present study identifies a multi-potent progenitor clone through characterization of its progeny and suggests that cell plasticity may be responsible for the observed heterogeneity in cell behavior.

TABLE OF CONTENTS

INTRODUCTION

While most human tumors are thought to be derived from a single cell, tumor cell populations demonstrate considerable heterogeneity in phenotype and behavior. It is increasingly appreciated that this intratumoral heterogeneity is central to metastasis, progression of disease and treatment failure. A recent brainstorming meeting between leaders in the field recognized "identifying the source of heterogeneity in cancer" as one of the most fundamental questions to address regarding tumor heterogeneity(1). Two theories have sought to account for the variation in cell phenotype and function: the cancer stem cell hypothesis and the clonal evolution model (2). The cancer stem cell model posits that a subpopulation of cells (found at the apex of the tumor hierarchy) retains tumor initiating abilities capable of reconstituting the entire cellular heterogeneity of the tumor while accumulating genetic alterations (3,4). Universal cancer stem cell (CSC) traits include an unlimited self-renewal potential through symmetric cell division as well as the ability to generate non-CSC (differentiated) progeny through asymmetric division(3). On the other hand, the clonal evolution model proposes that over time, genetic mutations accumulate in cell populations and under environmental pressures and natural selection, clonal outgrowths dictate the phenotype of the tumor (2,5,6). Studies have shown that some cancers follow a CSC model, whereas others are best explained by the clonal evolution model(7). Yet, some argue that the two theories need not be exclusive and rather an intermediate model, incorporating principles from both may better account for phenotypic heterogeneity(8).

In many solid malignancies, including ovarian cancer, genomic heterogeneity within the primary tumor sample and between metastatic sites has been extensively documented (9–13). Nevertheless, few studies have sought to correlate this genomic variability with functional characterization of cell subpopulations(8,14). Understanding how genetic and epigenetic differences influence cell behavior is essential to identifying the sources of intratumoral heterogeneity. Furthermore, a better knowledge of this heterogeneity can be exploited in designing more effective therapies, especially for metastatic and recurrent tumors. A recent study in glioblastoma was one of the first to isolate single cells from patient tumors and compare their genetic heterogeneity with functional readouts(14). Nevertheless, the single cell derived clones were cultured under stem cell conditions biasing their study toward more tumorigenic subpopulations.

In establishing an experimental model for characterizing the genomic and functional differences between primary tumor clonal populations, we utilized a new culture system developed by Tan Ince that has been shown to retain the genomic landscape, histopathologic and molecular features of original tumors(15). We utilized a primary clear cell ovarian cancer line (OCI-C5x) established by Ince to generate single-cell derived clonal populations in a non-biased manner. During their characterization, we identified a clone (Clone 31) with high tumor initiating activity when injected into the peritoneum of immunocompromised mice compared to the other single-cell clones which were either incompetent to expand *in vivo* or did so after a very long latency. Although very aggressive, Clone 31 did not form any solid metastasis, but rather proliferated in the peritoneum as ascites. Therefore, Clone 31 was not able to recapitulate the full functional heterogeneity of the original tumor. Its parent (OCI-C5x) on the other hand, formed solid metastases that resemble clear cell ovarian cancer histology. Nevertheless, Clone 31 was an exceptionally tumorigenic clone that warranted further characterization to better understand its functional and genetic traits.

Morphologically Clone 31 demonstrated a mixture of cells with epithelial and mesenchymal traits, suggesting that the original single cell clone was a multi-potent progenitor. We thus wanted to determine whether tumor initiating activity and other properties of these cells were restricted to progenitors, epithelial-like or mesenchymal-like cells. To accomplish this, we generated single-cell subclones of Clone 31 using fluorescence-activated cell sorting. In this study, we describe the characteristics of Clone 31's progeny in an attempt to elucidate the role of Clone 31 as a multi-potent progenitor. Single-cell subclones were subjected to a variety of functional *in vitro* and *in vivo* studies. Furthermore, genomic and expression differences were explored and correlated with subclone behavior and morphology.

STUDENT ROLE

Teodora Kolarova was involved with study design, experiment performance and data analysis. Specifically, she performed nearly all of the *in vitro* experiments including monolayer proliferation, soft agar colony formation and general cell culture maintenance. She designed and implemented the first *in vivo* experiment, which was repeated by Elaine Lunsford. She analyzed all of the *in vitro* and *in vivo* data. Lastly, she isolated genomic DNA and total RNA from the

subclones. Sequencing was carried out through our collaborators (Nick Navin and Gordon Mills) and data analysis of the genomic DNA and RNA sequencing was performed by Laura Selfors.

METHODS & MATERIALS

Reagents, Cell Culture

All cells used in this study were cultured in OCMI medium (LTCC, University of Miami). Cell line OCI-C5x was graciously provided by Tan Ince, MD, PhD (University of Miami). Clones and subclones were generated using fluorescence-assisted cell sorting (FACSAria cell sorter; BD Biosciences, Inc.) Single-cells were sorted into a 96 well plate under sterile conditions and expanded in OCMI media. Prior to the generation of the subclones, Clone 31 was infected with a vector expressing tdTomato fluorescent protein for visual detection and *Gaussia* Luciferase, a secreted luciferase. Vector was a gift from our collaborators. The vector was transduced via a lentivirus, generated by transfection of 293T cells. Lentivirus-infected cells were selected for 72 hours in medium containing 1ug/ml of puromycin (Dulbecco). All subclones inherited their vector from Clone 31.

Flow Cytometry

Cells were cultured in OCMI medium and harvested with Trypsin-EDTA during log-phase growth. Cells were fixed in 4% paraformaldehyde and permeabilized with 1:10 methanol. Blocking was performed with 5% goat serum in PBS. Primary antibodies used were mouse anti-E-Cadherin at 1:1000 dilution (BD Biosciences; Cat No 610181) and chicken anti-Vimentin (BioLegend; Cat No PCK-594P) at 1:1000 dilutions. Secondary antibodies used were Alexa Fluor 647 goat anti-mouse IgG antibody (BioLegend; Cat No 405321) and Alexa Fluor 488 goat anti-chicken IgY antibody (AbCam: Cat No Ab150169). Samples were analyzed at the Harvard Medical School Flow Cytometry Core Facility on a FACSCalibur (BD Biosciences, Inc). Experiments were carried out in triplicate. Flow cytometry data was analyzed using FlowJo v7.6.5 (FlowJo LLC) and GraphPad Prism (GraphPad Software, Inc.)

Two-Dimensional Monolayer Growth Rate

To establish *in vitro* proliferation rates, each subclone was plated into a 12-well tissue culture plate at a density of $7x10^3$ cells. Cells were harvested by trypsinization on day 1 and day

5 and counted using a particle counter (Z1; Beckman Coulter, Inc.). Experiments were carried out in triplicates. Doubling time was calculated with the assumption of exponential growth. The number of generations =$(log^{Day\ 5\ count}_{Day1\ count})/(log2)$. Doubling time was derived by dividing the time by the number of generations.

Soft Agar Colony Formation

To study anchorage-independent colony formation, a soft agar assay was conducted. A 0.5% agar dilution (Difco Agar Nobel) was plated in 6 well plates and allowed to solidify at 4ºC. Cells from each subclone were seeded in triplets of $4x10^4$ and mixed into a 0.4% agar top layer. Soft agar plates were stored at 37°C, 5% CO_2 for 21 days and refed with 0.4% agar media every 4 days. Colonies were stained with Neutral Red solution (Sigma-Aldrich) and subsequently imaged with a dissecting microscope. ImageJ was utilized to count & analyze particles. The soft agar assay was carried out in triplicate for each subclone. Data was analyzed GraphPad Prism (GraphPad, Inc.). An unpaired Students t Test was used to compare number of colonies between the subclones. The Mann Whitney test was used to compare median values of colony size.

***In Vivo* Xenograft Experiments**

Care of the mice was in accordance with the Harvard University's Institutional Animal Care and Use Committee (IACUC). Mice were maintained and handled under aseptic conditions, and animals were allowed access to food and water *ad libitum*. Female NOD/SCID mice were purchased from the Jackson Laboratory (Bar Harbor, Maine) at an age between 6 and 10 weeks.

Three million cells per subclone were harvested and injected into the peritoneum of a female NOD/SCID mouse. Three mice per subclone were used for each experiment and the experiment was repeated twice. The secreted *Gaussia* Luciferase (G-Luc) was utilized to monitor tumor growth(16). Blood was collected every two weeks from the submandibular vein of each mouse. G-Luc activity was measured after addition of coelenterazine using a plate luminometer (MLX luminometer, Dynex technologies, Chantilly, VA). The mice were euthanized 10 weeks after tumor cell injection. Upon death, each mouse was dissected and visual inspected for solid tumor metastasis. Ascites fluid was quantified as well.

Fold-change in G-Luc activity was calculated by standardizing to Day 1 G-Luc levels (day after injection). G-Luc fold-change from each mouse were averaged for each time point and

graphed on a log scale using GraphPad Prism. Individual time point statistical analysis was carried using the Students t Test.

Copy Number Analysis

Genomic DNA was extracted from each subclone using a cell culture DNA extraction kit (Qiagen, Inc.). DNA was quantified using the BioTek microplate reader (BioTek Instruments, Inc.). Two micrograms of genomic DNA was next sonicated into 200bp segments using the Covaris focus acoustics system. Next-generation sequencing libraries were prepared and barcoded using a standard Illumina library preparation protocol with end repair, 3' A-overhang addition and barcode ligation. Barcodes utilized were collection designed and verified by the Nick Navin laboratory. Library purification was undertaken using the AMPure bead purification system (Agencourt Bioscience Corporation). Next, libraries were amplified and quantified using the Agilent Bioanalyzer instrument. Finally, the samples were pooled and multiplexed. Sequencing was carried out using the Illumina HiSeq 2000 instrument at 10 million reads per sample (1-3x coverage). Data was segmented using Circular Binary Segmenter (CBS) and copy number profiles were calculated at 200kb resolution. Detailed protocol of copy number analysis using next generation sequencing has been previously described(17).

Deep Exome Sequencing

Sequencing was undertaken using T200, a deep targeted sequencing platform developed by the Institute for Personalized Cancer Medicine (IPCM) at MD Anderson Cancer Center (Houston, TX)(18). Two hundred and two genes were selected for deep exome sequencing. These genes were selected based on mutation prevalence, "actionability" (drug in trial or in the pipeline) and input from experts in the field(18).

Genomic DNA was extracted from each subclone using a cell culture DNA extraction kit (Qiagen, Inc.). DNA was quantified using the BioTek microplate reader and quality was accessed using Genomic DNA Tape for the 2200 Tapestation (Agilent). DNA from each sample (170-500 ng of genomic DNA) was sheared by sonication using Covaris E220 instrument (Covaris). To ensure the proper fragment size, samples were checked on TapeStation using the DNA High Sensitivity kit (Agilent). The sheared DNA proceeded to library prep using KAPA library prep kit (KAPA) following the "with beads" manufacturer protocol. Purification was

undertaken using Agencourt AMPure PCR Purification kit (Agencourt Bioscience Corporation). At the end of the library prep, samples were analyzed on TapeStation to verify correct fragment size and to ensure the absence of extra bands. Samples were quantified using KAPA qPCR quantification kit. Equimolar amounts of DNA were pooled for capture.

Biotin labeled DNA probes were designed by the IPCM (MD Anderson) with Roche Nimblegen for capturing target regions (over 5,000 exons in 202 genes) and followed manufacture's protocol for the capture process. The capture process and amplification have been previously summarized. The captured libraries were sequenced on an Illumina HiSeq 2000 (Illumina Inc., San Diego, CA, USA). The results were demultiplexed using CASAVA 1.8.2 with no mismatches. All regions were covered by >20 reads. Sequencing customized to ensure detection of low frequency (5%) variants. A custom data analysis pipeline was used to analyze data as previously described. To understand the potential functional consequence of detected variants, they were compared with dbSNP, COSMIC, 1000 Genomes, and TCGA databases. Further, the variants were annotated using SIFT, Polyphen, Condel, Mutation Assessor.

RNA Sequencing

RNA sequencing was carried out at the Biopolymers Facility at Harvard Medical School (Boston, MA). Total RNA was extracted from each subclone using the RNeasy Mini Kit according to manufacturer's protocol (Qiagen, Inc.). PolyA-tailed mRNA was selected using beads with oligodeoxythymidine. Quantity and quality of mRNA were assessed using BioAnalyzer Pico Chip (Agilent Technologies).

First, mRNA was reverse transcribed and purified using the PrepX mRNA Library protocol and the Apollo Wafergen 324 System (IntegenX Inc.). The purified cDNA was PCR amplified for 15 cycles and validated for quality and concentration using a BioAnalyzer (Agilent). Subsequently, the cDNA fragments went through end repair, end adenylation and barcode adaptor ligation according to manufacturer's protocol (Illumina Inc., San Diego, CA). After completion of the cDNA library prep, samples were analyzed on AgilentTape Station High-Sensitivity Tape to verify correct fragment size. Samples were quantified using KAPA qPCR quantification kit. Equimolar amounts of cDNA were pooled for capture and sequenced Illumina HiSeq2000 instrument (Illumina Inc., San Diego, CA).

Low quality sequence was eliminated from paired end reads with trimmomatic-0.3. The resulting reads were aligned to human genome assembly hg19 by bowtie2. Count data was generated by HTSeq-0.5.4 and subjected to scaling factor normalization in edgeR. Differential expression between two negative subclones (P1D16 and P2G21) and the positive subclones was performed in edgeR. Data were centered on the mean of the negative clones and log transformed. Hierarchical clustering was performed in Cluster 3.0 (Average linkage, uncentered) and visualized in JavaTreeView 1.1.1.

RESULTS

Generation and Morphologic Characterization of the Subclones

Sub-clonal populations were generated by sorting single cells from Clone 31 using FACS (BD Biosciences) and allowing the single cells to proliferate under the same *in vitro* conditions as the parent (Clone 31). Of the single cells able to reconstitute a new cell line, all were cultured continuously (up to 20 passages) with no decrease in growth rate.

When examined by phase-contrast microscopy, the eight subclonal populations exhibit a variety of cell morphologies from purely epithelial to mesenchymal-like to a mix of the two (Fig 1). Cells classified as epithelial were polygonal in shape and tended to grow in discrete patches. On the other hand, mesenchymal-like cells resembled fibroblast in their spindle shape and lack of close cell-cell contact. Clone 31 and the parental (OCI-C5x) are classified as mixed morphology. Each subclone maintained its specific morphology during passaging.

To further study this heterogeneity in morphology, we quantified the vimentin-positive and E-cadherin-positive populations within each subclone using flow cytometry. E-cadherin is a membrane glycoprotein involved in cell-cell adhesion and a marker of epithelial cells. Vimentin is an intermediate filament and a major component of mesenchymal cell structure. Loss of E-cadherin and expression of vimentin are hallmarks of the epithelial-to-mesenchymal transition, which leads to phenotypic changes responsible for differences in the properties of tumor cells, e.g. motility and invasiveness (19). P2E22 and P2L16 exhibit the highest proportion of vimentin-positive cells with 22.6% (SD 0.87) and 14.6% (SD 0.88), although the two subclones are of mixed morphology. Somewhat surprisingly, P2G21, a subclone of mostly epithelial-like morphology also has 14.1% (SD 4.08) vimentin-positive cells. The two most mesenchymal-like subclones, P1D16 and P2G16, have the lowest amount of vimentin positive cells at only 2.0%

(SD 0.36) and 0.8% (SD 0.31), respectively. Thus, morphologic appearance *in vitro* did not necessarily correlate with the epithelial and mesenchymal biomarkers.

In Vitro Activity of Subclones

Next, we characterized the heterogeneity *in vitro* growth properties of the subclones. First, we determined doubling time, which ranged from 26.8 (SD±3.4) hours for P1P17 to 62.8 (SD±21.8) for P1D16 (Figure 2). OCI-C5x had a doubling time of 22.0 (SD±2.1) and Clone 31 had a doubling time of 29.1 (SD±2.4). Doubling time was not associated with cell morphology ($p>0.05$). Overall, the doubling time of the subclones was not statistically different compared to Clone 31's.

Furthermore, we utilized a soft agar colony formation assay to test the ability of the subclones to grow in an anchorage-independent manner. Colony formation independent of a solid scaffold is a hallmark of cell transformation and carcinogenesis. The subclones demonstrated differences in both their ability to proliferate in soft agar and in the distribution of colony sizes (Figure 2). For example, subclones P2L16, P1C14 and P1P17 formed similar number of colonies as their parent (Clone 31). On the other hand, P2G16, P2G21 and P1D16 formed the least amount of colonies and relative to Clone 31, these differences were statistically significant. Regarding colony size, all subclones exhibited statistically significant reductions in colony median size compared to Clone 31 with the exception of P2L16 (Figure 2). Ability to form colonies in soft agar did not correlate with *in vitro* morphology. However, both P1D16 and P2G16 showed the lowest proportion of Vimentin positive cells and also formed significantly lower number of colonies relative to their parent. Therefore, anchorage-independent growth was not uniform among the subclones with some demonstrating lesser colony formation compared to their parent.

In Vivo Tumor Initiating Activity of the Subclones

Since the subclones demonstrated considerably varied morphology and diverse *in vitro* behavior, we sought to determine each subclones' ability to generate tumors *in vivo*. As mentioned previously, Clone 31 was of particular interest because of its tumorigenic abilities *in vivo* and its distinct phenotype of ascites without any solid metastasis. In contrast, the parental bulk line, OCI-C5x, forms ascites and solid metastasis on the mesentery, liver, ovary and

diaphragm, which recapitulate the ovarian clear cell histology. Therefore, as a representation of Clone 31's progeny, the *in vivo* phenotype of the subclones was also of interest.

To monitor tumor growth, Clone 31 was transfected with a vector containing Gaussia Luciferase (G-Luc), a secreted luciferase. G-Luc allows for the monitoring of tumor progression in real-time via blood draws and/or digital imaging(16). The transfection was performed prior to the generation of the subclones and thus all subclones contain the same vector as their parent. Tumor growth was calculated as the fold change in blood G-Luc levels compared to the Day 1 (post-injection) levels. Three million cells of each subclone were injected into the peritoneal cavity of female NOD/SCID mice. Blood was collected from each mouse every other week. Mice were euthanized at 10 weeks post-injection and underwent autopsy to characterize phenotype.

At 10 weeks, the majority of subclones, including P1C14, P1M14, P2E22, P1P17 and P2G16, showed similar growth patterns compared to Clone 31 (Figure 3). Interestingly, P2L16 demonstrated more growth at 10 weeks than its parental line ($p = 0.03$). P1D16 and P2G21 were the least tumorigenic when compared to Clone 31 ($p=0.0006$ and 0.001, respectively).

Another objective of the *in vivo* experiment was to observe whether the subclones remained restricted to ascites growth only like their parent or if they gained an ability to form solid metastasis. Of the tumorigenic subclones, P1P17 and P2G16 did not form solid tumors, just like their parent clone (Figure 3). On the other hand, P1M14, P2E22, P2L16 and P1C14 all formed solid tumors in locations commonly associated with ovarian cancer metastasis including the ovary, omentum, liver and diaphragm. Comparison of the *in vivo* activity with *in vitro* morphology and soft agar colony formation did not reveal a clear correlation. Interestingly, only subclones of mixed two-dimensional morphology were capable of forming solid metastasis.

Copy Number Variation

Copy number variations (CNVs) are an important source of diversity among normal physiologic processes as well as pathologic ones (17,20). CNVs describe deletions or duplications within the genome of at least 1 kb in size. To study these, we utilized whole-genome next-generation sequencing, which is becoming the preferred method of interrogating CNVs due to its widespread availability and quality of data (17,21).

All the subclones remained diploid with the exception of P2G16, which showed an amplification of chromosome 15. Compared to their parent, the subclones did not exhibit additional copy number variations (Figure 4). However, compared to OCI-C5x, Clone 31 and its progeny did demonstrate an amplification of *ERBB2.* These findings are consistent with recent work showing that *ERBB2* amplification (determined by IHC and FISH) is more common in clear cell ovarian cancer samples compared to other subtypes of ovarian cancer (22).

Mutation Analysis via Deep Targeted Exome Sequencing

Although no major copy number variations were detected in the subclones compared to Clone 31, we wanted to ascertain whether other genomic alterations could account for the heterogeneity of *in vivo* tumorigenicity among the subclones. In addition, OCI-C5x was known to harbor mutations in *MSH2* (R711*) and *ATM* (Y731C). In general, *MSH2* nonsense mutations have been implicated in a variety of malignancies due to faulty mismatch repair (23). Interestingly, endometrioid and clear cell ovarian cancers are more likely to exhibit DNA mismatch repair deficiency than the epithelial counterparts (23). With this knowledge, we had observed that single-cell clones from OCI-C5x, Clone 31 included, possessed many unique mutations indicating that this tumor had a "mutator phenotype." Further, this raised the possibility that Clone 31 could undergo additional genetic alterations during *in vitro* passage. Subsequently, each subclone was subjected to deep targeted exome sequencing of a panel of 202 genes selected based on mutation prevalence and "actionability" (18). To focus on particularly relevant mutations, we specifically evaluated those amino acid substitutions that were predicted to affect protein functionality.

Three mutations are shared among Clone 31 and its subclones that are not found in OCI-C5x (Table 1). Four other mutations are present in a lower frequency in Clone 31, but enriched in some of the subclones. These mutations include *SMARCA4* R1256H, *KDR* A739T, *AKT1* G162S and *SPTA1* Q700*. The *SMARCA4* variant is present in a very low frequency in Clone 31 indicating that it is most likely a new mutation occurring after cloning. Conversely, the *KDR*, *AKT1* and *SPTA1* variants are present at a higher frequency (about 0.25 on average) and increase to a 0.50 frequency in select subclones. This is surprising given maintenance of diploid status across the subclones with P2G16 being the exception. Interestingly, the *KDR* A729T mutation is prevalent in the subclones unable to proliferate *in vivo* (P1D16, P2G21) and form solid tumors

(P2G16, P1P17), the only exception being P2L16 which retains this mutation but is able to proliferate and form solid metastasis. Subclone-specific mutations are also present with the exception of P2E22 and P1C14. Most variants are presents at 0.40-0.50 frequency which suggests that they were present at the time of sorting, but at a low enough frequency not to be detected in Clone 31. The subclone-specific variants present at low frequencies are most likely new mutations that occurred after the initial cloning. Given the *MSH2* and *ATM* mutation, subclone-specific mutations were not surprising.

Gene Expression Pattern of the Subclones

To further elucidate the differences between the subclones, we sought to determine if any expression signatures were correlated with tumorigenic potential *in vivo*. Gene expression was studied using RNA sequencing (Illumina HiSeq2500 Platform). Data was normalized to the average of P1D16 and P2G21, both of which were unable to proliferate *in vivo*.

One-hundred and thirteen genes were differently expressed between the two "negative" subclones and the rest of the subclone panel (Figure 5). To analyze whether these genes were enriched for specific pathways or classes of genes, we utilized the Molecular Signature Database (MSigDB v5.1) and GeneGo MetaCore (Thomson Reuters) (24). The genes whose expression was upregulated in the tumorigenic subclones were enriched for genes involved in "SLC-mediated transmembrane transport" and in the "transport of glucose, other sugars, bile salts, organic acids metal ions and amine compounds". These twenty-two genes also showed a statistically significant likelihood of being transcriptional targets of CREB1 and SP1 ($p=1.25e^{-26}$ and $p=6.04e^{-20}$, respectively). On the other hand, the genes, whose expression was downregulated in the tumorigenic subclones, were enriched for genes involved in extra-cellular matrix and cytoskeleton remodeling. Furthermore, twenty-two of these ninety-one genes overlapped with a set of genes found to be down-regulated after *CDH1* (E-Cadherin) knockdown (FDR $p=7.92e^{-14}$). The transcriptional factor analysis showed this set to be enriched for targets of Androgen Receptor, CREB1, ESR1, p53 and cMyc.

DISCUSSION

The present study was designed to address the role of a tumorigenic single-cell clone as a potential multi-potent progenitor by characterizing its progeny. The study also attempted to

elucidate mechanisms of heterogeneity among the subclones. These experiments provide evidence for the plasticity of tumor cells based on functional and genomic heterogeneity among the eight single-cell subpopulations. This diversity is summarized in Table 2.

While Clone 31 displayed a strong ability to initiate tumor growth *in vivo,* this property was not universally retained among the subclones. Furthermore, a fraction of these subclones were capable of forming solid tumors, an attribute that Clone 31 does not possess. While our study was designed to identify processes and properties that correlated with tumorigenicity, we were unable to distinguish any that clearly correlated with this activity. Neither cell morphology, nor colony formation, nor vimentin expression was distinctly associated with *in vivo* activity; however, it is noteworthy that only subclones with mixed epithelial and mesenchymal 2D morphology were able to form solid metastasis. It has been suggested that epithelial-mesenchymal plasticity (EMP), which encompasses EMT and the mesenchymal-epithelial transition (MET), is necessary for successful metastasis (25). It is possible that the subclones with mixed morphology also experience the greatest plasticity under appropriate microenvironmental cues, enabling them to form solid tumors. A static capture of E-cadherin and vimentin levels under monolayer culturing conditions may not capture this malleability and thus may explain why this data did not correlate with tumorigenicity in the immunocompromised mouse model. EMP also implies the existence of an intermediate population co-expressing both mesenchymal and epithelial markers. In fact, several recent papers present findings that challenge the concept of a strict epithelial and mesenchymal dichotomy (26,27). In their study, Grosse-Wilde et al. identified a hybrid epithelial/mesenchymal breast cancer cell population that showed the most plasticity, self-renewal and mammosphere formation, all markers of stemness (26). In addition, co-expression of E and M signatures was correlated with worse overall survival for all breast cancer subgroups except Luminal B tumors(26). Consistent with this, our flow cytometry data supports the existence of a dually stained population, which showed variability from trial to trial (data not shown). Additionally, RNA Sequencing of an EMT gene set demonstrated hybrid expression in the subclones capable of proliferating *in vivo*, whereas the negative subclones cluster with the non-tumorigenic clones (Supplemental Figure 1). Although suggestive of a co-expression state and EMP, further studies need to validate these findings in the subclones. Furthermore, the implication of epithelial-mesenchymal plasticity in successful metastasis is a novel, emerging development in the context of the CSC paradigm. Classically, the

model identified CSCs as a small, biologically unique subset of cells capable of perpetuating tumor growth indefinitely in a hierarchical manner(28). However, new evidence suggests that CSC characteristics are not innate to one subpopulation of cells, but rather dynamically gained or lost based on cell-cell, stromal or microenvironmental interactions (29,30). Furthermore, although the epithelial-mesenchymal transition has been implicated in the "stemness" of both normal and cancerous cells, evolving data supports the notion that EM plasticity rather than EMT alone is critical for CSC-driven tumor expansion (19,30–32).

The gene expression data of the subclones also corroborates a role for epithelial-mesenchymal plasticity in functional diversity. For instance, a good proportion of the genes differentially expressed in the tumorigenic subclones are implicated in metabolism, cytoskeletal remodeling and E-cadherin knockdown, all key pathways affected by EMT & MET(25,33–35). Interestingly, one of the genes differential expressed was *FOXC2*. *FOXC2* expression was lower in the subclone that expanded in the peritoneal cavity as ascites or solid tumors. *FOXC2* overexpression has been associated with EMT (36–38). Since EMT has been so fervently correlated with invasion and metastasis, our findings seem contradictory. However, a recent study has shown that MET and E-cadherin are required for ovarian cancer cells to form spheroids and survive when injected in the peritoneal cavity (39). Perhaps the two subclones expressing higher levels of *FOXC2* are "stuck" in an expression milieu that prevents them from reverting back to a phenotype necessary for their survival and growth in the peritoneal cavity of the mice. It has been shown that overexpression of transcription factors implicated in EMT such as Snail and Twist, leads to decreased metastatic potential(40,41). Therefore, epithelial-mesenchymal plasticity, rather than EMT markers may be more predictive of tumorigenicity. Another possibility is that higher *FOXC2* expression does not signify EMT in these subclones, especially since no other EMT markers were concurrently expressed. A major limitation of the gene expression data is its source: cells cultured in a monolayer. Although under standardized conditions, in the absence of microenvironmental cues and stromal interactions, the cellular pathways responsible for *in vivo* tumorigenicity may not be readily evident. These cellular programs are dynamic and may only be activated or suppressed once within the peritoneal cavity of the mouse, which provides a starkly different environment than a culture dish. Changes in gene expression are essential for understanding the survival, proliferation and metastasis of these subclones and warrant further thought.

Of the factors contributing to intratumoral heterogeneity, the most widely studied are genomic alterations. Genomic instability is a hallmark of cancer attributed in part to the destabilization of gene copy number and in part to the large number of DNA repair defects documented in human cancer (42). Instability in OCI-C5x is evident by the *MSH2* and *ATM* mutations as wells as the large number of unique variants in this tumor. Most of the somatic mutations that are exclusive to the subclones have not been studied or reported in the literature, which makes elucidating their role difficult. Nevertheless, a few warrant special mention and further investigation. As previously mentioned, all four subclones incapable of forming solid metastasis harbor the *KDR* A739T mutation. *KDR* encodes vascular endothelial growth factor receptor 2 (VEGFR2), a receptor tyrosine kinase that is activated upon VEGF ligand binding and is a key player in cell adhesion, migration and angiogenesis (43). Although this mutation is predicted to be deleterious, it has not been reported previously. Residue 739 lies in the proximal-most extracellular component of the protein. Mutations at residues 726 and 731 have been reported to disrupt salt bridges between VEGFR2 homodimers, which are necessary for auto phosphorylation (44). If this mutation impacts the subclones' ability to adhere and recruit angiogenic components, it may provide additional explanation as to why they do not proliferate or metastasize in the mouse. Interestingly, P2L16 also harbors this mutation, but forms robust solid tumors. However, P2L16 has a mutation in *SMARCA4*, a core component of the multi-subunit SWI/SNF chromatin remodeling complex. This mutation has been previously reported in the undifferentiated component of a dedifferentiated endometrial carcinoma (45). The tumorigenic phenotype of P2L16 in the context of the VEGFR2 mutation may be explained by malfunctioning SWI/SNF and resultant widespread epigenetic dysregulation. Lastly, although the novel subclone alterations may be explained by aberrant mismatch repair, it is unclear if they were present in a very small proportion (<1%) of cells prior to sorting or if they occurred subsequent to sorting during culturing.

The sources of intratumoral heterogeneity are numerous and dynamic. Genetics, epigenetics and the microenvironment are all implicated in cancer cell diversification. The present study provides evidence that Clone 31 forms a functionally diverse progeny consistent with a multi-potent progenitor. While we were unable to identify any processes or properties that accounted for tumorigenicity in the subclones, observations from our study warrant further investigation into epithelial-mesenchymal plasticity as a significant source for this heterogeneity.

ACKNOWLEDGEMENTS

I want to express my gratitude to Joan Brugge, PhD for mentoring me through my project and giving me the opportunity to work with many of her talented doctoral and post-doctoral students. Further recognition goes to Suha Naffar-Abu Amara whose expertise, training and everyday support were invaluable to this project. Thank you to Elizabeth, Carina and Kripa who carried on the project during my clinical years. Also, we are all very grateful for the generosity in time and commitment of our collaborators including Tan Ince, Nick Navin, Marco Leung and Gordon Mills. Last, but definitely not least, a large thank you to Laura Selfors for her bioinformatics expertise.

REFERENCES

1. Alizadeh AA, Aranda V, Bardelli A, Blanpain C, Bock C, Borowski C, et al. Toward understanding and exploiting tumor heterogeneity. Nat Med. Nature Publishing Group; 2015;21(8):1–8.
2. Campbell LL, Polyak K. Breast Tumor Heterogeneity: Cancer Stem Cells or Clonal Evolution? Cell Cycle. 2007;6(19):2332–8.
3. Baccelli I, Trumpp A. The evolving concept of cancer and metastasis stem cells. J Cell Biol. 2012 Aug 6;198(3):281–93.
4. Curley MD, Garrett LA, Schorge JO, Foster R, Rueda BR. Evidence for cancer stem cells contributing to the pathogenesis of ovarian cancer. Front Biosci. 2011 Jan;16(2):368–92.
5. Greaves M, Maley CC. Clonal evolution in cancer. Nature. 2012;481(7381):306–13.
6. Marjanovic ND, Weinberg R a, Chaffer CL. Cell Plasticity and Heterogeneity in Cancer. Clin Chem. 2012 Dec 6;179.
7. Shackleton M, Quintana E, Fearon ER, Morrison SJ. Heterogeneity in Cancer: Cancer Stem Cells versus Clonal Evolution. Cell. 2009;138(5):822–9.
8. Calbo J, van Montfort E, Proost N, van Drunen E, Beverloo HB, Meuwissen R, et al. A Functional Role for Tumor Cell Heterogeneity in a Mouse Model of Small Cell Lung Cancer. Cancer Cell. Elsevier Inc.; 2011;19(2):244–56.
9. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012 Mar 8;366(10):883–92.
10. Khalique L, Ayhan A, Whittaker JC, Singh N, Jacobs IJ, Gayther SA, et al. The clonal evolution of metastases from primary serous epithelial ovarian cancers. Int J Cancer. 2009;124(7):1579–86.
11. Schwarz RF, Ng CKY, Cooke SL, Newman S, Temple J, Piskorz AM, et al. Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. PLoS Med. 2015 Mar;12(2):e1001789.
12. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. Nature. 2011 Apr 7;472(7341):90–4.
13. Kim T-M, Jung S-H, An CH, Lee SH, Baek I-P, Kim MS, et al. Subclonal Genomic Architectures of Primary and Metastatic Colorectal Cancer Based on Intratumoral Genetic

Heterogeneity. Clin Cancer Res. 2015 Oct 1;21(19):4461–72.

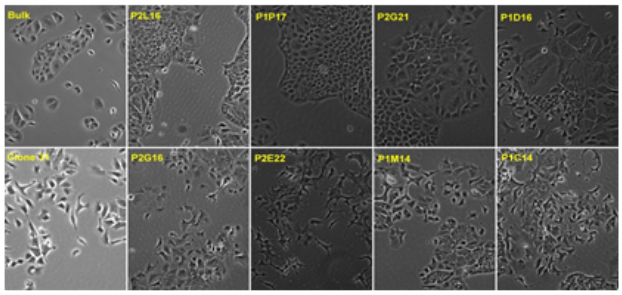
14. Meyer M, Reimand J, Lan X, Head R, Zhu X, Kushida M, et al. Single cell-derived clonal analysis of human glioblastoma links functional and genomic heterogeneity. Proc Natl Acad Sci U S A. 2015;112(3):851–6.
15. Ince TA, Sousa AD, Jones MA, Harrell JC, Agoston ES, Krohn M, et al. Characterization of twenty-five ovarian tumour cell lines that phenocopy primary tumours. Nat Commun. Nature Publishing Group; 2015;6(May):7419.
16. Chung E, Yamashita H, Au P, Tannous B a, Fukumura D, Jain RK. Secreted Gaussia luciferase as a biomarker for monitoring tumor progression and treatment response of systemic metastases. PLoS One. 2009 Jan;4(12):e8316.
17. Baslan T, Kendall J, Rodgers L, Cox H, Riggs M, Stepansky A, et al. Genome-wide copy number analysis of single cells. Nat Protoc. Nature Publishing Group; 2012;7(6):1024–41.
18. Chen K, Meric-Bernstam F, Zhao H, Zhang Q, Ezzeddine N, Tang L-Y, et al. Clinical actionability enhanced through deep targeted sequencing of solid tumors. Clin Chem. 2015 Mar;61(3):544–53.
19. Polyak K, Weinberg RA. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. Nat Rev Cancer. 2009;9(4):265–73.
20. Li W, Olivier M. Current analysis platforms and methods for detecting copy number variation. Physiol Genomics. 2013 Jan 7;45(1):1–16.
21. Liu B, Morrison CD, Johnson CS, Trump DL, Qin M, Conroy JC, et al. Computational methods for detecting copy number variations in cancer genome using next generation sequencing: principles and challenges. Oncotarget. 2013 Nov;4(11):1868–81.
22. Zannoni GF, Morassi F, Prisco MG, De Stefano I, Vellone VG, Arena V, et al. Clinicopathologic and immunohistochemical features of ovarian clear cell carcinomas in comparison with type I and type II tumors. Int J Gynecol Pathol. 2012 Nov;31(6):507–16.
23. Xiao X, Melton DW, Gourley C. Mismatch repair deficiency in ovarian cancer -- molecular characteristics and clinical implications. Gynecol Oncol. 2014 Feb;132(2):506–12.
24. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545–50.

25. Ye X, Weinberg RA. Epithelial-Mesenchymal Plasticity: A Central Regulator of Cancer Progression. Trends Cell Biol. Elsevier Ltd; 2015;25(11):675–86.
26. Grosse-Wilde A, Fouquier d'Hérouël A, McIntosh E, Ertaylan G, Skupin A, Kuestner RE, et al. Stemness of the hybrid Epithelial/Mesenchymal State in Breast Cancer and Its Association with Poor Survival. PLoS One. 2015 Jan;10(5):e0126522.
27. Sampson VB, David JM, Puig I, Patil PU, de Herreros AG, Thomas G V, et al. Wilms' tumor protein induces an epithelial-mesenchymal hybrid differentiation state in clear cell renal cell carcinoma. PLoS One. 2014 Jan;9(7):e102041.
28. Valent P, Bonnet D, De Maria R, Lapidot T, Copland M, Melo J V, et al. Cancer stem cell definitions and terminology: the devil is in the details. Nat Rev Cancer. Nature Publishing Group; 2012 Nov;12(11):767–75.
29. Bomken S, Fiser K, Heidenreich O, Vormoor J. Understanding the cancer stem cell. Br J Cancer. 2010 Aug 10;103(4):439–45.
30. Luo M, Brooks M, Wicha MS. Epithelial-mesenchymal plasticity of breast cancer stem cells: implications for metastasis and therapeutic resistance. Curr Pharm Des. 2015 Jan;21(10):1301–10.
31. Foster R, Buckanovich RJ, Rueda BR. Ovarian cancer stem cells: Working towards the root of stemness. Cancer Lett. 2012 Nov 5;
32. Scheel C, Weinberg R a. Phenotypic plasticity and epithelial-mesenchymal transitions in cancer and normal stem cells? Int J Cancer. 2011 Nov 15;129(10):2310–4.
33. Cha YH, Yook JI, Kim HS, Kim NH. Catabolic metabolism during cancer EMT. Arch Pharm Res. 2015;38(3):313–20.
34. Nieto MA. Epithelial plasticity: a common theme in embryonic and cancer cells. Science (80-). 2013;342(6159):1234850.
35. Kalluri R, Weinberg R a. Review series The basics of epithelial-mesenchymal transition. J Clin Invest. 2009;119(6):1420–8.
36. Liu B, Han S-M, Tang X-Y, Han L, Li C-Z. Overexpressed FOXC2 in ovarian cancer enhances the epithelial-to-mesenchymal transition and invasion of ovarian cancer cells. Oncol Rep. 2014 Jun;31(6):2545–54.
37. Li Q, Wu J, Wei P, Xu Y, Zhuo C, Wang Y, et al. Overexpression of forkhead Box C2 promotes tumor metastasis and indicates poor prognosis in colon cancer via regulating

epithelial-mesenchymal transition. Am J Cancer Res. 2015 Jan;5(6):2022–34.

38. Cui Y-M, Jiao H-L, Ye Y-P, Chen C-M, Wang J-X, Tang N, et al. FOXC2 promotes colorectal cancer metastasis by directly targeting MET. Oncogene. 2015 Aug 13;34(33):4379–90.
39. Elloul S, Vaksman O, Stavnes HT, Trope CG, Davidson B, Reich R. Mesenchymal-to-epithelial transition determinants as characteristics of ovarian carcinoma effusions. Clin Exp Metastasis. 2010;27(3):161–72.
40. Tsai JH, Donaher JL, Murphy DA, Chau S, Yang J. Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. Cancer Cell. 2012 Dec 11;22(6):725–36.
41. Celià-Terrassa T, Meca-Cortés O, Mateo F, Martínez de Paz A, Rubio N, Arnal-Estapé A, et al. Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. J Clin Invest. 2012 May;122(5):1849–68.
42. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. Elsevier; 2011 Mar 4;144(5):646–74.
43. Ferrara N. Vascular endothelial growth factor: basic science and clinical progress. Endocr Rev. 2004 Aug;25(4):581–611.
44. Yang Y, Xie P, Opatowsky Y, Schlessinger J. Direct contacts between extracellular membrane-proximal domains are required for VEGF receptor activation and cell signaling. Proc Natl Acad Sci U S A. 2010 Feb 2;107(5):1906–11.
45. Karnezis AN, Wang Y, Ramos P, Hendricks WPD, Oliva E, D'Angelo E, et al. Dual loss of the SWI/SNF complex ATPases SMARCA4/BRG1 and SMARCA2/BRM is highly sensitive and specific for small cell carcinoma of the ovary, hypercalcemic type. J Pathol. 2015;(August 2015):n/a – n/a.
46. Taube JH, Herschkowitz JI, Komurov K, Zhou AY, Gupta S, Yang J, et al. Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. Proc Natl Acad Sci U S A. 2010 Aug 31;107(35):15449–54.

A



Cell Line	Morphology
OCI-C5x (Bulk)	Mixed
Clone 31	Mixed
P1P17	Epithelial-like
P2G21	Epithelial-like
P2L16	Mixed
P1C14	Mixed
P1M14	Mixed
P2E22	Mixed
P1D16	Mesenchymal-like
P2G16	Mesenchymal-like

B

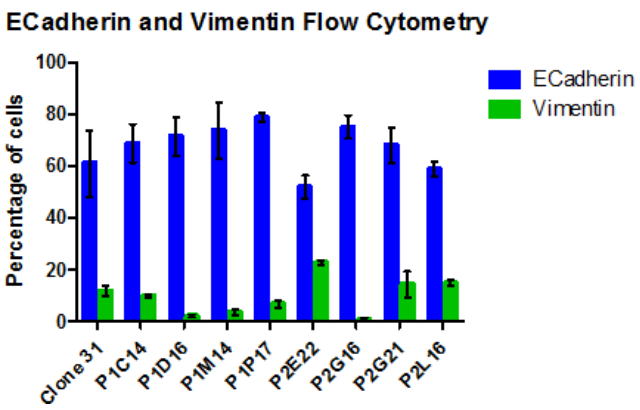


Figure 1: **Characterizing *In Vitro* Subclone Morphology**. (A) Representative phase contrast microscopy images of each subclone with a table summarizing cell morphology. (B) Quantification of E-Cadherin (blue) and Vimentin (green) expressing cell populations using flow cytometry. Data averaged over three replicate trials. Error bars represent standard deviation

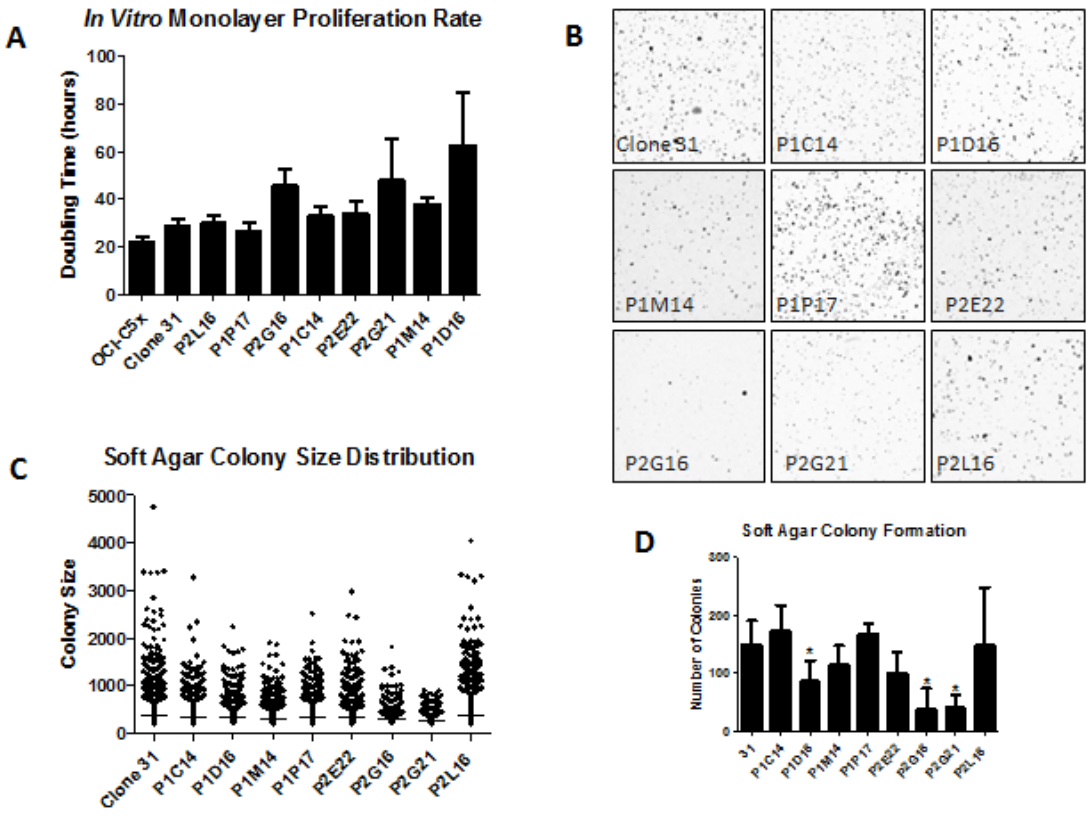
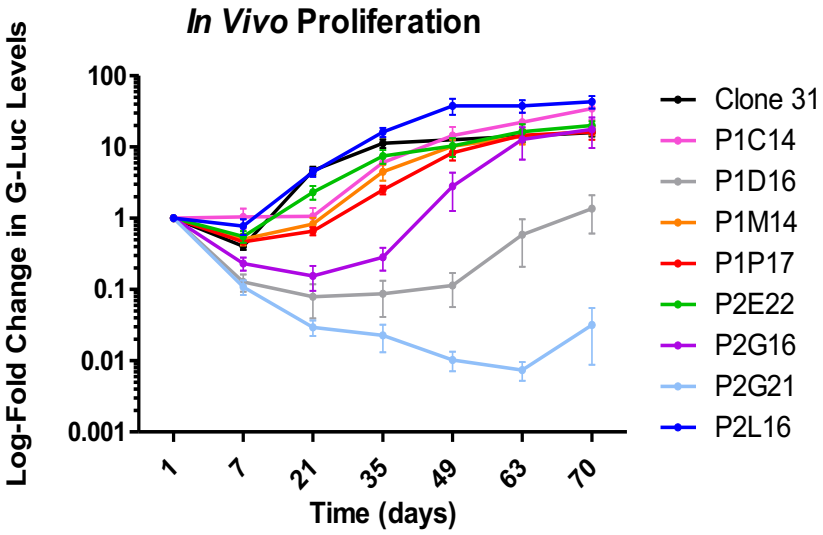


Figure 2: **Characterizing subclonal proliferation rate and colony formation** (A) Doubling time in hours under two-dimensional monolayer conditions. Data averaged over three trials with error bars representing standard deviation. No significant correlation between doubling time and morphology observed. (B) Representative images of colony formation in soft agar (1.5X magnification). Images processed and analyzed using ImageJ. (C) Depiction of colony size distribution for each subclone. Median values represented by horizontal lines. Compared to Clone 31, all subclones had significantly smaller median values ($p<0.05$) with the exception of P2L16 which showed a colony size distribution similar to Clone 31. (D) Quantification of the total number of colonies per well per subclone. Data averaged over two replicate experiments with error bars representing standard deviation. Asterisks represent a statistically significant difference in mean number of colonies compared to Clone 31.



Clone	PHENOTYPE	Ascites	Solid tumor
Clone 31	ASCITES	~1.5ml x6	no
P1P17	ASCITES	~1ml x3 ~0.5 ml x3	no
P2G16	ASCITES	~0.5ml x5	no
P1M14	MIXED	~1ml x6	diaphragm, liver (4/3)
P2E22	MIXED	~1.5ml x6	Ovary, omentum, liver
P2L16	MIXED	~1.5ml x5	ovary, omentum, liver (2/5)
P1C14	MIXED	~1.5ml x4 ~1ml x1 ~0.5ml x1	ovary, omentum
P1D16	NEG>	~50µl x3 ~ 10µl x2 no x1	no
P2G21	NEG>	no	no

Figure 3: **Characterizing subclone *in vivo* growth.** Three million cells per subclone were injected into the peritoneum of female NOD/SCID mice. Tumor growth was monitored via secreted Gaussia Luciferase (G-Luc) into the blood. Growth was characterized by fold-change of G-Luc activity compared to Day 1 levels. Data pooled from two replicate experiments (six mice total per subclone). At day 70, P1D16 and P2G21 showed significantly less growth compared to Clone 31 (Students t test: $p<0.05$). Table summarizes phenotype of each subclone within the mouse upon dissection at day 70.

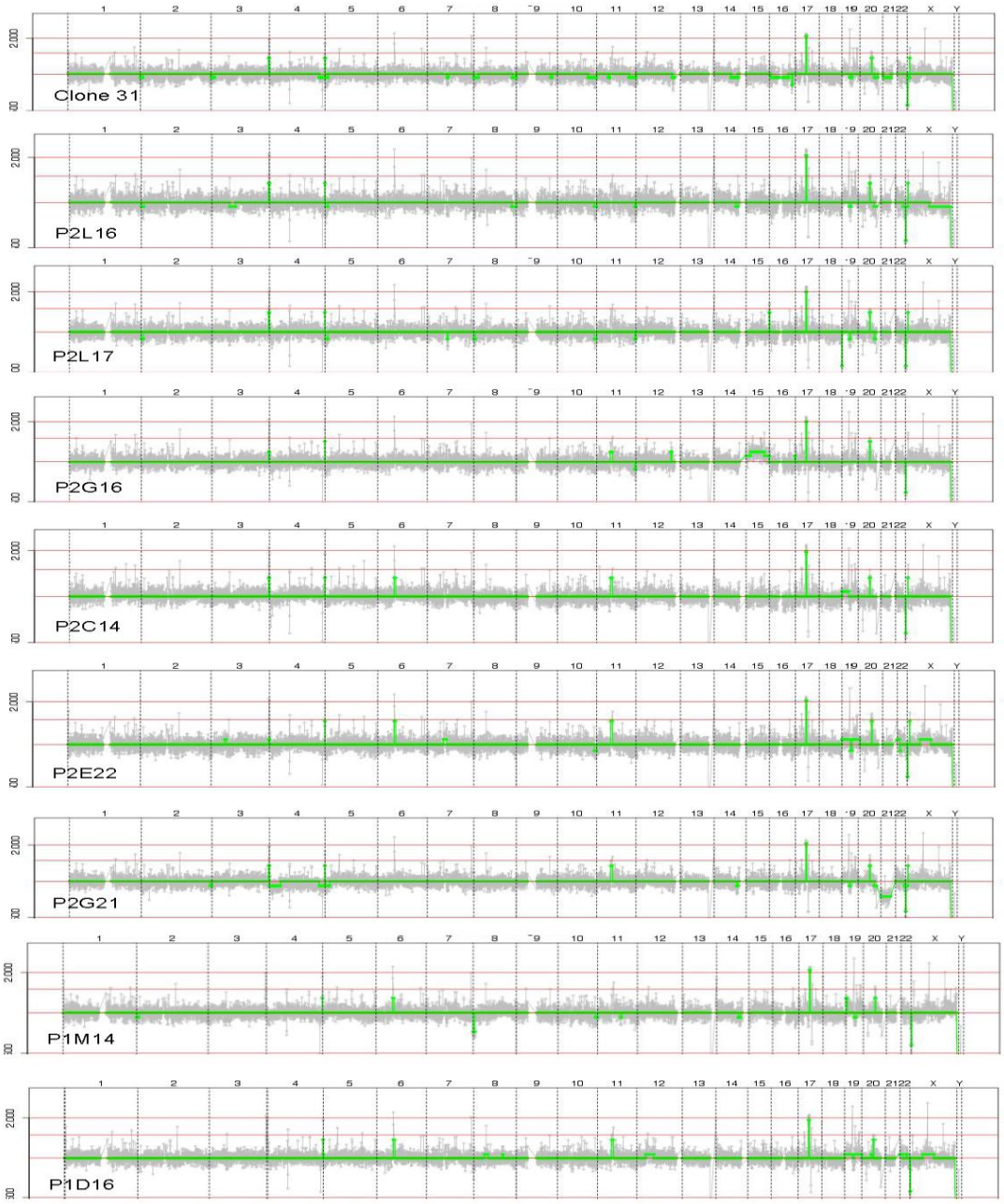


Figure 4: **Copy Number Analysis**. Copy number analysis undertaken using next generation sequencing of genomic DNA. P2G16 was the only subclone to demonstrate aneuploidy with an amplification of chromosome 15. All subclones including Clone 31 exhibit an amplification of a portion of chromosome 17 containing *ERRB2.*

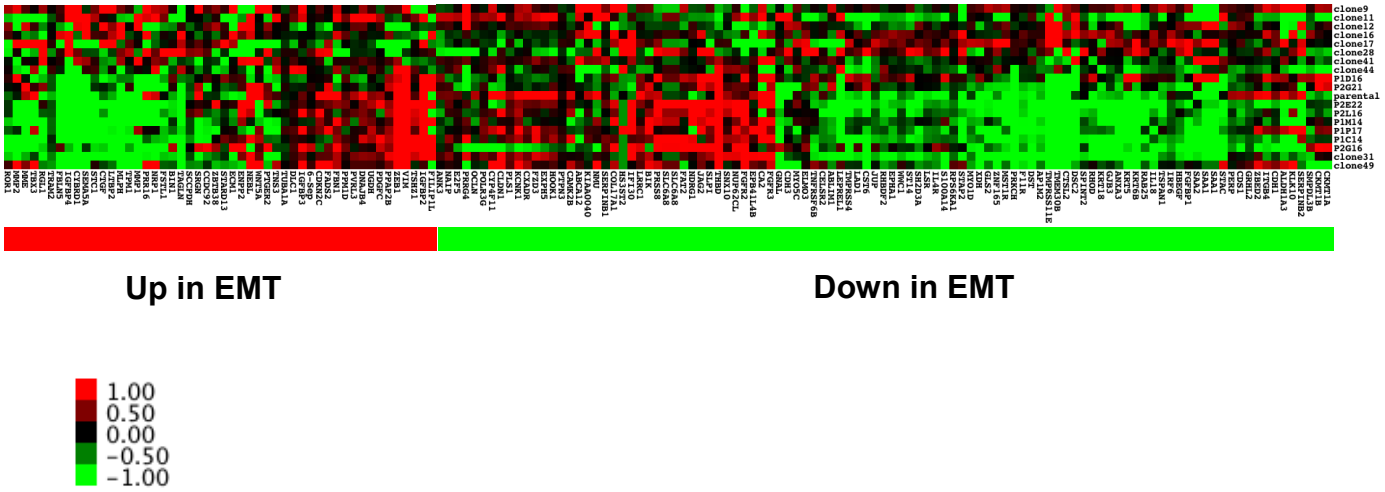
Gene (HGNC)	Amino Acid Substitution	Clone 31	P2L16	P2E22	P1C14	P1M14	P1P17	P2G16	P1D16	P2G21
In Vivo Phenotype		Positive, no solid tumor	Positive w/ solid tumors	Positive w/ solid tumors	Positive w/ solid tumors	Positive w/ solid tumors	Positive, no solid tumor	Positive, no solid tumor	Negative	Negative
CSMD3	G2284D	0.4338	0.476	0.4766	0.4799	0.448	0.4679	0.429	0.4828	0.4572
ERCC5	Q755*	0.463	0.4545	0.4632	0.503	0.4435	0.4734	0.5029	0.4966	0.4593
RNF213	D2900V	0.4857	0.4493	0.5155	0.4696	0.4922	0.506	0.487	0.4538	0.4882
SMARCA4	R1256H	0.0549	0.5208							
KDR	A739T	0.2254	0.5273				0.468	0.474	0.454	0.4819
AKT1	G162S	0.2583		0.4663	0.4688	0.4873				
SPTA1	Q700*	0.2621		0.5091	0.4861	0.503				
CSMD2	I3060F								0.3693	
CSMD2	I3060M								0.3665	
CSMD2	S3059F								0.3737	
MTOR	R2270W								0.379	
MET	C385W								0.4529	
MET	K380Q								0.4605	
ARAF	T442M								0.4847	
CSMD1	R709M					0.121				
SPEN	P2940L					0.2798				
CSMD1	C2720Y					0.4558				
RUNX1T1	D139G						0.4338			
PRDM1	P481S						0.4442			
SPTA1	H2282Q						0.4909			
GNAS	Q627*							0.1474		
MDN1	S2268N							0.1727		
IGF1R	Y973C							0.4396		
PRDM1	R477M							0.483		
CSF1R	D712N									0.4964
PCLO	P1758H		0.4897							

Table 1: **Deep exome sequencing of the subclones**. Table summarizes mutations unique to Clone 31 and its subclones. Numerical values represent the frequency of said mutation with a detection threshold of 5%. Only those variants predicted to be deleterious are shown. Mutations in blue are those that are not present in OCI-C5x, but are found in Clone 31 and the subclones. Mutations in red are found in Clone 31 and only some of the subclones. Gene variants in green are only found in one of the subclones, thus representing either accumulating mutations or mutations present in Clone 31 at less than 10% frequency.

Figure 5: **Gene expression analysis using RNA Sequencing.** Gene expression values were standardized to the mean of P1D16 and P2G21, the two subclones incapable of proliferating *in vivo*. Heatmap depicts the 113 genes differentially expressed between the two groups of subclones.

Cell Line	Morphology	Vimentin+ %	Soft Agar	*In Vivo*	Mutations
Clone 31	Mixed	11.6%	+++	Ascites	KDR
P1P17	Epithelial-like	6.6%	++	Ascites	KDR, RUNX1T1, PRDM1, SPTA1
P2G16	Mesenchymal-like	0.8%	-	Ascites	KDR, GNAS, MDN1, IGF1R, PRDM1
P2L16	Mixed	14.6%	+++	Ascites + solid tumor	KDR, SMARCA4, PCLO
P1M14	Mixed	3.3%	++	Ascites + solid tumor	AKT1, SPTA1, CSMD1, SPEN
P2E22	Mixed	22.6%	++	Ascites + solid tumor	AKT1, SPTA1
P1C14	Mixed	9.7%	++	Ascites + solid tumor	AKT1, SPTA1
P1D16	Mesenchymal-like	2.0%	+	Negative	KDR, CSMD2, MTOR, MET, ARAF
P2G21	Epithelial-like	14.1%	-	Negative	KDR, CSF1R

Table 2: **Summary table of subclone *in vitro* and *in vivo* acitivty**



Supplementary Figure 1: **Expression of an EMT gene signature among the subclones and clones of OCI-C5x.** The EMT core signature was derived from the changes in gene expression shared by up-regulation of Gsc, Snail, Twist, TGF-b1 and by down regulation of E-cadherin. (46). Data is median of negative centered. P1D15 and P2G21, the non-proliferative in vivo subclones cluster with the nontumorigenic clones.