Profiling of paired leukemia cell lines and primary tumors

Jeremiah Wala
Biophysics 205
April 26, 2013

Introduction and Motivation

Cancer cell lines are a ubiquitous pre-clinical model for testing drug sensitivity, genomic alterations and signaling pathways in many different tumor types. They are especially attractive because of the ease with which they can be manipulated *in vitro*, where finely controlled microenvironments can measure response to drugs, gene knockdown or culture conditions. More recently, there has been an interest in high-throughput characterization of the genomic alterations present in cancer cell lines. Notable projects include the Cancer Cell Line Encyclopedia (CCLE) (Barretina 2012) in the US and the Cancer Cell Line Project (Forbes 2011) in the UK. Additionally, a large scale RNAi screen against ~11000 genes has been recently applied to 102 cell lines to identify genes that are essential for cancer growth (Project Achilles), potentially identifying novel targets for drug development (Cheung 2011).

Despite efforts to characterize cancer cell lines, little is known about how the process of creating and passaging an immortalized cell line changes the underlying biology of the cells. Real tumors do not live in a culture dish, but rather reside in a complex *in vivo* environment. The surrounding stroma provides mechanical support, growth hormones and cytokines signal to tumor cells, and our own immune system constantly challenges the growing tumor. Additionally, the concentration of oxygen, nutrients and vasculature varies greatly across tumors, and may contribute to drug and radiation resistance (Gray 1953). Thus, cells in culture face an entirely different fitness landscape, and it follows that alterations selected for *in vivo* may confer no selective advantage in culture (or may be damaging). Further, mutations that are passengers (non-advantageous) *in vivo* may drive expansion of that clone in culture. Previous studies have tracked cytological changes in cell line over time (Macville 1999) and compared cell line transcriptomes to unmatched primary tumors (Gillet 2011). However, to my knowledge, there have been no genome wide studies to systematically identify genes that are uniquely advantageous to cancer cell culture, rather than for *in vivo* proliferation. The discovery of such alterations could redirect drug development towards more relevant genes, and would provide important information for designing future cell-line based experiments.

The relative fitness of cancer cells *in vitro* is expected to be governed by the acquisition of protein coding mutations, as well as by changes in gene regulation that alter gene expression. For this reason, I will quantify how the genome and transcriptome of cultured cell lines evolves after removal from the primary tumor. Three new cell lines will be derived from patients with acute myeloid leukemia (AML), and whole genome sequencing (WGS) and RNA-sequencing (RNA-seq) will be applied to matched primary and cell cultures at two times points: low passage cells (LPC) and high passage cells (HPC). I will interpret the results of sequencing in the context of data from the CCLE and Project Achilles to generate a list of genes that are significantly altered, or have significant expression changes, following the transition to cell culture.

Hematopoietic cancers provide the best system for this study for several reasons. Tumors are inherently heterogeneous, containing multiple subclones with unique genomic features. Ideally, one would want the established cell lines to derive from the same

location as the cells sent directly for sequencing, to minimize confounding from tumor heterogeneity. Because leukemia cells are not mechanically confined to surrounding stroma, they are well mixed -- cells sent for culture and cells sent for immediate sequencing will be drawn from the same population. Additionally, establishing leukemia cell lines is technically easier, as lines can be established directly from peripheral blood draws without having to be enzymatically dissociated or passaged through a xenograft. Finally, many leukemia patients (specifically AML) are young, and leukemia cancer genomes have low rates of mutations and simpler somatic genomes (Ley 2008).

Although previous cancer cell line sequencing projects do not sequence matched primaries, whole genome sequencing (WGS) of temporally and spatially separated tumors within a single patient are now being reported. Similar to the methods proposed here, Ding *et al* performed WGS on matched normal, primary, metastasis and xenograft samples from a single breast cancer (Ding 2011). By looking at mutations, structural variations and copy number changes across the four sequences, they successfully identified private mutations in the metastasis and xenograft, indicative of continuous tumor evolution. The approach proposed here is similar in methodology, but directed towards a different question: what are the genomic changes that provide fitness to cells *in vitro*?

Specific Aim 1: Establish three AML cancer cell lines from separate clinical samples

Heparinized peripheral blood will be obtained from pediatric patients being treated at the Dana-Farber Cancer Institute with a diagnosis of AML confirmed by a cytopathologist. The clinical characteristics of patients will depend on the patient population at the time of the study. Leukemia cell lines from both treated and treatment naïve patients have been previously established (Asou 1991, Sun 2004). Age, treatment status, and time to remission will be recorded for all patients. The success rate for establishing cell lines from peripheral blood in B-cell precursor cell lines is between 5% and 30%, with treatment naïve patients being the most difficult (Drexler 2011). To obtain three AML cell lines, I expect to attempt cultures for 10-20 patients with AML.

I will modify an established protocol for isolating leukemia cell lines from peripheral blood, based on previous successful reports of establishing AML cell lines (Sun 2004, Drexler 2011). Samples will be diluted 1:2 in DMEM and centrifuged using a Ficoll-Hypaque density gradient solution. The mononuclear layer at the top will be removed, washed, and separated into equal partitions. Each aliquot will be counted and tested for viability. One vial will be immediately frozen using RNAlater (Invitrogen) at -196^{o}C to preserve the RNA and DNA state of the cells. Following establishment of a cell line from the other aliquot, this sample can be thawed, and profiled as described in Aim 2 and Aim 3.

Cells for culture will be resuspended to 10^6 / mL and incubated in 10% CO_2 at 37^{o}C. Due to the difficulty of establishing cytokine-independent cell cultures, cell media will be supplemented with GM-CSF, IL-3, PIXY-321, SCF and IFN-gamma, which have been applied with high success rates to AML (Drexler 1997). Leukemia cells rapidly

proliferate in the first 4 weeks following isolation, and cells will be passaged every two days.

There are several potential difficulties in establishing long-term leukemia cell lines. One challenge is distinguishing tumor cells from Epstein-Barr Virus (EBV) transformed lymphoblast cells, which also immortalize and proliferate rapidly. After 2 weeks of culture, I will perform RT-PCR to look for copies of the EBV+ lymphoblasts. Additionally, I will test all cell lines for the presence of mycoplasma species, a common and damaging contamination frequently encountered when establishing novel cell lines.

After two months of continuous culture, I will freeze portions of the cell lines for later culture and sequencing, and these cells will be designated as the low-passage cells (LPC). After twelve months of continuous cultures, the cells will be designated as high-passage cells (HPC). The total number of passages during this time will be recorded to establish a per-passage event rate for the leukemia cell lines.

Specific Aim 2: Profile the genomic differences between primary tumors and cell lines

To find altered genes that confer an advantage specific to cell culture, I will perform whole genome sequencing at 60x coverage through the Genome Sequencing Platform at the Broad Institute from the cells generated in Aim 1. Additionally, to ensure that genomic alterations represent somatic mutations, matched normal tissue will be sequenced to 30x for each patient. Mutation calls for all samples will be identified using Genome Analysis Took Kit (GATK) software (Broad Institute). Structural rearrangement will be identified using the dRanger algorithm, which identifies paired-end reads that map to distant genomic locations (Berger 2011). The purity and ploidy of each sample will also be identified using the ABSOLUTE algorithm, a method that uses SNP chip data to estimate the heterogeneity of tumors (Carter 2012).

Mutations and structural rearrangements present in all three samples will be considered to represent true alterations, following manual review of the relevant reads (as in Ding 2011). Potentially functional alterations (mutations and small indels present in coding regions, splice sites, non-coding RNA genes and within 1 kb of transcription start sites) present in either the LPC or HPC cells will be will be validated by PCR amplification and sequencing. Alterations not present in the primary will be called *private* alterations, which may have occurred *de novo* or from growth of a previously undetected subclone. Additionally, all cells will be karyotyped to identify large-scale translocations, a frequent driver event in hematopoietic cancers.

Due to the higher rates of genomic alterations found in cancer cells and cell lines, I expect many private alterations to arise during the course of cell culture. In this case, with only three samples it will be challenging to identify individual genes and alterations that confer advantages specific to the *in vitro* environment. To overcome this challenge, I will narrow the list of candidate genes using data from the 192 hematopoietic cell lines profiled in the Cancer Cell Line Encyclopedia (CCLE), as outline in Figure 1. Because

CCLE performed targeted sequencing rather than WGS, current methods for determining mutation significance will inapplicable (e.g. InVex, Hodis 2012). Instead, I will create an algorithm that identifies genes that are statistically more likely to mutated than expected by chance by permuting mutations within samples, controlling for gene size, overall mutation rate, and GC content. Significantly altered genes are expected to have either contributed to oncogenesis, proliferation in cell culture, or both. Additionally, to identify genes that are frequently amplified or deleted, I will apply the GISTIC algorithm (Mermel 2010) to the SNP data from the 192 CCLE leukemia cell lines. Briefly, GISTIC uses SNP probe intensity across many samples to find regions of the genome with frequent copy number alterations.

Genes that have been mutated, deleted, amplified or rearranged during the process of cell culture will be cross-referenced to the list of significantly altered genes across the 192 CCLE cell lines, creating a list of candidate genes that are advantageous for cell growth in culture. To validate that these alterations provide a growth advantage, I will apply shRNAs against the altered genes and compare the growth curve against a sham shRNA control. For each gene, three different shRNAs will be applied, and knockdown will be confirmed by western blot. In addition, I will apply the shRNAs to the six leukemia cell lines from the NCI-60 panel. In the case that the gene has a known inhibitor, I will create drug sensitivity curves for the three novel cell lines and the six NCI-60 leukemia lines.

Specific Aim 3: Profile the transcriptome in the primary tumors and cell lines

I will perform RNA-seq on the primary tumors and LPC and HPC cells to quantify RNA expression levels among the samples. Sequencing reads will be aligned, normalized and quantified with GenePattern (Broad Institute). To establish significance values for differential expression, the Cuffdiff (Trapnell 2013) algorithm will be applied to each sample independently, across all three time points. Briefly, Cuffdiff uses variability in read depth along a gene, along with mapping quality scores, to provide an estimated variance of gene expression. Cuffdiff will then compare expression among the conditions (i.e. time points) and provide a ranking of genes that are differentially expressed.

As with identifying significantly mutated genes, the actual amount of variation in gene expression between primary and culture cells is unknown. With only three samples, one challenge that could arise is that there may be insufficient power to identify genes whose expression is more significantly altered than others. To narrow the list of differentially expressed genes to ones that have the greatest biological significance to growth in cell culture, I will utilize data from the Achilles project, a large shRNA screen of 102 cell lines that measured the relative abundance of cells following gene knockdown. Genes that are upregulated specifically in novel cell lines should be expected to confer a growth and survival advantage to the cells in culture, and their knockdown in Achilles should inhibit cell growth. Conversely, genes that are downregulated should have either no effect on cell survival in culture, or inhibit cell growth *in vitro*. The major drawback to this analysis is that among the 102 cell lines, only two are hematopoietic (multiple myeloma and leukemia).

Finally, to understand the biological significance of the differentially expressed genes, I will use GenePattern to perform gene ontology (GO) enrichment analysis to identify pathways that are significantly altered in cell culture. Additionally, I will perform two-dimensional hierarchical clustering on the gene expression profiles to identify if cells are more similar among cases within a time point (i.e. culture time is most important indicator of similarity), or among different time points within a case. To further understand gene expression changes, I will repeat the clustering after incorporating gene expression profiles from 40 additional leukemia cell lines (Andersson 2005).

Additional Possible Experiment

Although not included in my specific proposal, a recently reported data set of WGS for eight matched primary and recurrent post-treatment AML tumors suggests an additional interesting experiment (Ding 2013). They find novel mutations and a high incidence of transversions in AML cells following chemotherapy. Ultimately, one would like to know how well cell line models recapitulate true tumor response to chemotherapy. Using the LPC and HPC cell lines, I would apply identical chemotherapy drugs as used in treatment, and resequence following the development of resistance to identify new mutations and rearrangements. This would address the question of whether the mechanism of drug resistance in cell lines is similar to the mechanism now seen in resistant *in vivo* tumors.

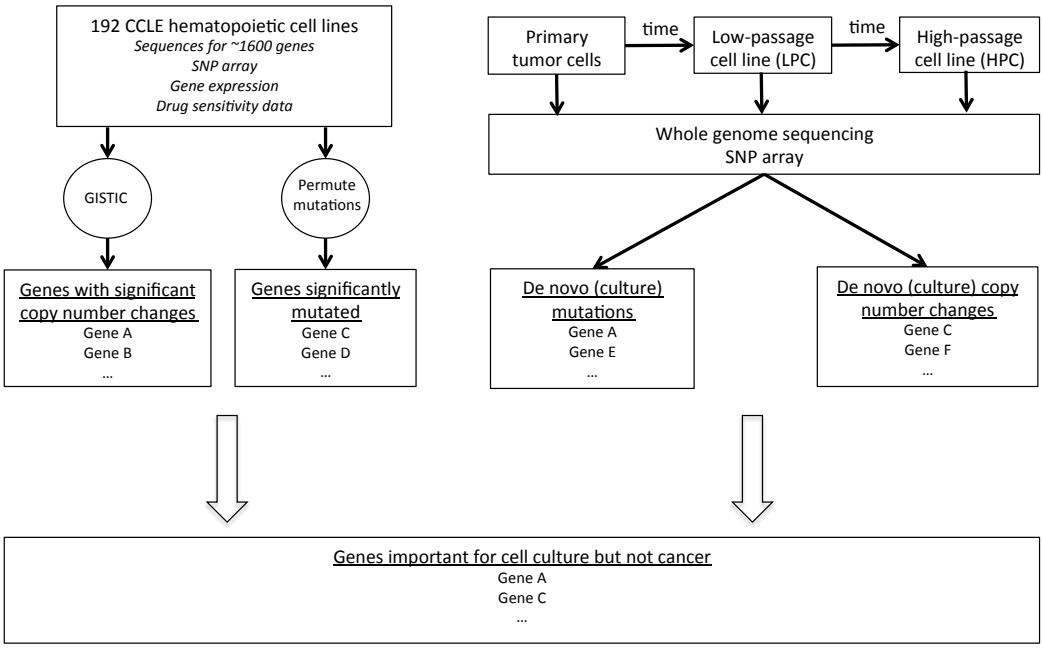

Figure 1: Flowchart of the proposed workflow to identify genes that are putatively important for cell culture, but not found to be altered in primary tumors. Because it is unknown what the initial mutation rate of leukemia cell lines is following cell culture establishment, incorporating CCLE data (as shown on left hand side) might be required to narrow the list of altered genes to those that are most promising.

References

Andersson, A., Edén, P., Lindgren, D., Nilsson, J., Lassen, C., Heldrup, J., et al. (2005). Gene expression profiling of leukemic cell lines reveals conserved molecular signatures among subtypes with specific genetic aberrations. Leukemia, 19(6), 1042–1050.

Asou, H., Tashiro, S., Hamamoto, K., Otsuji, A., Kita, K., & Kamada, N. (1991). Establishment of a human acute myeloid leukemia cell line (Kasumi-1) with 8;21 chromosome translocation. Blood, 77(9), 2031–2036.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature, 483(7391), 603–307.

Berger, M. F., Lawrence, M. S., Demichelis, F., Drier, Y., Cibulskis, K., Sivachenko, A. Y., et al. (2012). The genomic complexity of primary human prostate cancer. Nature, 470(7333), 214–220.

Carter, S., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. Nature Biotechnology, 30(5), 413–421.

Cheung, H., Cowley, G., Weir, B., Boehm, J., Rusin, S., Scott, J., et al. (2011). Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. PNAS, 108(30), 12372–12377.

Ding, L., Ellis, M. J., Li, S., Larson, D. E., Chen, K., Wallis, J. W., et al. (2011). Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature, 464(7291), 999–1005.

Drexler, H., Zaborski M., Quentmeier, H. (1997). Cytokine response profiles of human myeloid factor-dependent leukemia cell lines. Leukemia. 11(5);701—8.

Drexler, H. Establishment and culture of leukemia-lymphoma cell lines. (2011). Methods in Molecular Biology, Chapter 16. Humana Press.

Forbes, S., Bindal, N., Bamford, S., Cole, C., Kok, C., Beare, D., et al. (2010). COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Research, 39; D945–D950.

Gillet, J., Calcagno, A., Varma, S., Marino, M., Green, L., Vora, I., et al. (2011). Redefining the relevance of established cancer cell lines to the study of mechanisms of clinical anti-cancer drug resistance. PNAS, 108(46), 18708–18713.

Gray, L., Conger, A., Ebert, M., Hornsey, S., (1953). British Journal of Radiology. 26, 638–648.

Hodis, E., Watson, I. R., Kryukov, G. V., Arold, S. T., Imielinski, M., Theurillat, J.-P., et al. (2012). A landscape of driver mutations in melanoma. Cell, 150(2), 251–263.

Ley, T. J., Mardis, E. R., Ding, L., Fulton, B., McLellan, M. D., Chen, K., et al. (2008). DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature, 456(7218), 66–72.

Macville, M., Schröck, E., Padilla-Nash, H., Keck, C., Ghadimi, B. M., Zimonjic, D., et al. (1999). Comprehensive and definitive molecular cytogenetic characterization of HeLa cells by spectral karyotyping. Cancer Research, 59(1), 141–150.

Mermel, C., Schumacher, S., Hill, B., Meyerson, M., Beroukhim, R., & Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biology, 12(4), R41.

Sun, Y., Kim, S. H., Zhou, D.-C., Ding, W., Paietta, E., Guidez, F., et al. (2004). Acute promyelocytic leukemia cell line AP-1060 established as a cytokine-dependent culture from a patient clinically resistant to all-trans retinoic acid and arsenic trioxide. Leukemia, 18(7), 1258–1269.

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature Biotechnology, 31(1), 46–53.