Chapter 1

Introduction

1.1 Complexity of human cancers

Our current understanding of the molecular basis of cancer has been shaped by initial landmark studies demonstrating that cancer was caused by the deregulation of a few oncogenes or tumor suppressor genes (1, 2). Identification of these genes and their encoded protein products has had a tremendous impact on molecular medicine, leading to effective targeted therapies for certain cancers. For example, patients with chronic myelogenous leukemia (CML), a bone marrow derived cancer, express the fusion oncogene BCR-ABL that is effectively targeted by the kinase inhibitor imatinib mesylate (3–7). In breast cancer, elevated levels of the membrane-bound receptor oncogene Her2/neu (ERBB2) predict response to the humanized monoclonal antibody therapeutic trastuzumab (8, 9). The clinical effectiveness of targeted molecular therapies against certain types of cancer is a testament to the progress in cancer research and medicine.

It is becoming increasingly clear, however, that for the majority of metastatic cancers, a single or few genes is insufficient in predicting tumor behavior and does not represent all necessary and sufficient targets for molecular therapy. Tumors are highly complex, dynamic, heterogeneous biological systems driven by series of genomic and epigenetic mutations that affect the genes controlling cell proliferation, survival, invasiveness and motility. Typically, multiple primary genomic alterations are present within a single tumor with secondary mutations arising through genetic instability, enabling malignancies to develop resistance to drugs. For example, the majority of lung adenocarcinomas that respond to epidermal growth factor receptor (EGFR) inhibitors (erlotinib and gefitinib) eventually develop resistance through selection of tumor cells expressing an EGFR variant containing a single amino acid mutation that confers drug resistance (10–12). Pre-malignant tissue evolve in a multi-step process into tumors, each step highly sensitive to numerous interacting extracellular stimuli, including growth factors (13, 14), extracellular matrix proteases (15, 16), physiological stresses (e.g., low oxygen levels (17, 18)) and cytokines (19). Many sources contribute to the tumor milieu, including non-tumor, stromal cells found in the local architecture (20, 21) and immune cells that have infiltrated the tumor microenvironment (19, 22, 23). Thus for a typical malignancy, a large number of genomic, proteomic, physiological conditions, and supporting cell types are involved.

1.2 Diagnostic challenges presented by cancer

One of the goals of cancer diagnostics is to detect and monitor biological elements that are representative of the interactions between the various components involved in a malignancy, and with the information be able to characterize the state of the system and/or generate network models with predictive features that would be of fundamental or therapeutic value (**Figure 1.1**). The biological elements targeted are

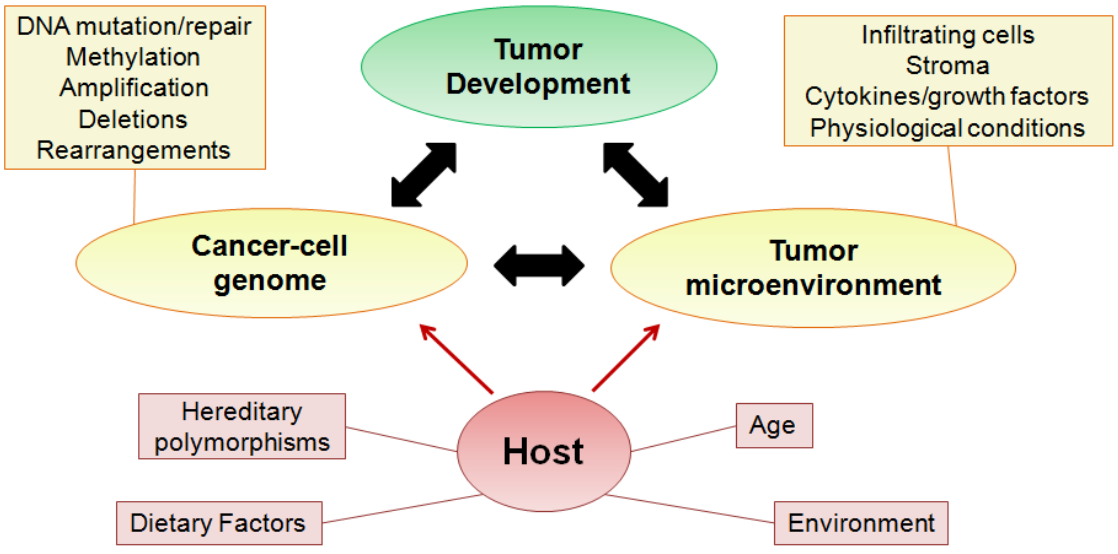

Figure 1.1 Numerous components of tumor biology must be queried for informative diagnostics. Reciprocal interactions between the tumor microenvironment, tumor cell genome and tumor development drive the complexity of cancer, presenting challenges for multi-parameter diagnostics. Adapted from (36).

either directly involved in the malignancy (e.g. over expression of EGFR), or are byproducts of the process (e.g. circulating tumor DNA resulting from necrosis). Individually, they are referred to as biomarkers. Currently, most FDA approved cancer diagnostic platforms focus on the detection of a single or few biomarkers and have been effective for several cancers for staging, monitoring, and prognostication (24). This pauci-parameter approach, however, is being revised because of the molecular

heterogeneity of cancer—the accuracy and coverage of diagnostic assays will be improved by adopting a multi-parameter systems approach, namely measure as many different biological elements as possible. The motivation is that a collection of biomarkers would produce a molecular signature that would provide a higher level of sensitivity and specificity for staging, treatment and prognostication.

A number of studies reported in the literature have shown that global profiling approaches are effective in characterizing cancer. Differential gene expression studies have shown that diseased complex expression networks can be differentiated from their healthy counterparts and can be used to predict patient survival and response to cancer therapies (25–28). Many of the proteins, genes and small molecules identified in these studies have spear-headed further investigations, including mechanistic studies (29) and the use of these targets as potential biomarkers (30–32). Other global profiling approaches investigating genomic alterations (33–34) and proteomic alterations (30, 35) have also been reported, although for the latter, studies have been impeded by the lack of proteome wide capture agents and lack of technologies to capture all the proteomic functions concomitantly (36) (e.g. interactions involving protein-protein, protein-ligand, post-translational modifications).

The adoption of a multi-parameter paradigm is compounded by the fact that most biological tissues collected from patients are small, making it challenging to execute large-scale analyses of the different classes of biological molecules (e.g. cells, genes, proteins, metabolites). This has driven the miniaturization of many diagnostic assays into integrated biochips. The benefits of developing biochips are severalfold. First, integrated biochips are able to handle small amounts of tissue. This is important in

interfacing with tumor tissue derived from skinny-needle biopsies and other minimally invasive diagnostic procedures. Second, on-chip measurements can be highly parallel and multiplexed which is important in providing statistical certainty by repeated sampling. Third, chip-based devices can be manufactured with well developed processes borrowed from mature engineering disciplines (e.g. semiconductor fabrication) and are typically compatible with a host of common instrumentation (e.g. fluorescent microscopy), lowering costs and increasing accessibility.

An example of a cancer diagnostic that has benefited from miniaturization is circulating tumor cell (CTC) detection. CTCs are cells shed from primary tumors and are in circulation at low levels in blood. The abundance of CTCs can be used to monitor the efficacy of cancer therapeutics. Current technologies for CTC detection rely on bead-based assays (37) but are plagued with low recoveries and low purity. Recently, two chip-based, microfluidic devices have been reported which significantly increase recoveries and purities of CTC capture (38, 39). The robustness of one biochip enabled the correlation of CTCs levels with decreased or increased tumor burden and the identification of mutations through genomic sequencing of CTCs collected from the device (40).

1.3 DNA as an encoding element

The utility of DNA as a chemical material has reached into many areas of active research, including nanomaterials (41–43), DNA computing (44), automated machines (45), and molecular electronics (46). In the arena of biological sensing involving mainly protein detection, technologies and platforms that have been developed include immuno-

PCR (47), rolling circle amplification (48), proximity ligation (49), and nanoparticle-based assays (50). DNA-based, multiplexed assays in which multiple detection events are encoded and then decoded simultaenously include immunophenotyping using transcription (51) and self-assembled chemical libraries (52).

There are several reasons that makes DNA an attractive material and encoding element. First, the exquisite specificity and favorable energetics of DNA base-pairing provides self-assembling properties, permiting complex 2-D and 3-D structures to be designed and predicted from linear primary sequences. Second, analagous to its endogenous function, unique DNA sequences can be used to store information, scaling exponentially to the length N of the polymer. Third, in its natural environment, DNA interfaces with a host of other biological elements which can be exploited for molecular level control. For example, DNA can be specifically truncated with restriction endonucleases. Fourth, the production of DNA oligonucleotides is trivial, owing in large part to chemical automation. A wide variety of unnatural chemical handles (e.g. primary amines, thiols), readouts (e.g. fluorophores), and modified bases (e.g. locked nucleic acids) are available for incorporation into a primary sequence, expanding the utility and applications of the polymer. Lastly, platforms to characterize DNA are robust, including sequencing, PCR, microarrays, and bioinformatic algorithms.

1.4 Thesis overview

This thesis presents the development of chip-based, DNA-encoded technologies to address the current multi-parameter challenges associated with *in vitro* cancer diagnostics. In Chapter 2, I will begin by introducing the development of an approach

called DNA-Encoded Antibody Libaries (DEAL) in which computationally derived, orthogonal ssDNA tags are conjugated to an antibody library where every antibody-specificity is uniquely encoded with a distinct ssDNA sequence. A library of DEAL conjugates is exposed to a biological sample, bind to their cognate antigen, and decoded spatially on a glass substrate printed with the complementary DNA sequences. I demonstrate the DEAL technique for; (1) the rapid detection of multiple proteins within a single microfluidic channel, and with the additional step of electroless amplification of gold-nanoparticle labeled secondary antibodies, establish a detection limit of 10 fM for the protein IL-2; (2) the multiplexed, on-chip sorting of both immortalized cell lines and primary immune cells; and (3) the co-detection of ssDNAs, proteins, and cell populations on the same platform. By using a common assembly molecule, DEAL meets the diagnostic need for multi-parameter platforms able to manipulate and detect major subtypes of biological molecules (genes, proteins, cell membrane-bound markers). Moreover, this technique is fully integrable with fluidics, enabling the processing of small tissue samples. Chapter 2 has been taken in part from © *J. Am. Chem. Soc*. **2007**, *129(7)*, 1959–67.

Chapter 3 extends and develops the concept of cell sorting with DNA tags in the context of an important immunological problem with clinical implications for cancer therapy, namely that of detecting antigen-specific T cells. Through genetic recombination, T cells are capable of recognizing and engaging to cells presenting processed antigenic fragments, including antigens presented by cancer cells. The ability of T cells to cull target populations of cells has been exploited as an experimental cancer therapy by infusing patients with augmented and activated T cells specific for cancer

associated antigens (53) with documented T cell-induced tumor regression in subsets of patients with metastatic cancers (54, 55).

Despite these promising trials, T cellular immunotherapy is difficult to characterize fully because of technological bottlenecks that do not allow the multi-parametric analysis of different antigen-specific T cells in small numbers. With the goal of highly multiplexed T cell detection, Chapter 3 will introduce technique called "Nucleic Acid Cell Sorting (NACS)" in which single-stranded DNA oligomers conjugated site-specifically to p/MHC tetramers are employed to immobilize p/MHC tetramers via hybridization to a complementary-printed substrate. Fully assembled p/MHC arrays are used to detect and enumerate T cells captured from cellular suspensions, including primary human T cells collected from cancer patients. Importantly, T cell array binding is optimized by utilizing cysteine-engineered streptavidin (SAC) for ssDNA-p/MHC tetramer production, resulting in NACS p/MHC arrays that outperform conventional spotted arrays assessed by performance standards such as reproducibility and homogeneity. The versatility of using DNA tags is also exploited to enable selective detachment of T cells with restriction endonucleases. Demonstrative experiments regarding NACS sensitivity, multiplexing and limit of detection are performed with cell lines and with T cells isolated from cancer patients. Lastly, I show an important clinical application of this technology by monitoring the presence and abundance of a cancer-specific T cell population collected from a melanoma patient receiving cellular immunotherapy. Chapter 3 has been taken in part from © *J. Am. Chem. Soc.*, **2009**, *in press*.

Cell surface proteins constitute an important subset of the cellular proteome, as these proteins are frequent targeted for molecular therapy, staging and for directing therapies. Chapter 4 presents current work on detecting cell membrane-bound proteins by amplification of encoded ssDNA tags. In this approach, DEAL/NACS conjugates are synthesized with photolabile ssDNA tags. Whereas in previous demonstrations, the ssDNA tags were employed as molecular tethers, in the cellular barcoding approach, the ssDNA tags are used as reporter molecules. After staining a population of cells with a library of DEAL/NACS conjuates, the ssDNA tags are released into solution via UV cleavage and collected for analysis by PCR. The quantity of each unique ssDNA tag directly correlates to the presence and expression of the targeted cell surface markers. I demonstrate that this concept by detecting EGFR over-expression from a low-passage brain cancer cell line (GBM1600) relative to EGFR null Jurkat human T cells. In addition, I show that different TCR specificities can be differentiated by this technique by using NACS conjugates to detect the presence of a TCR specific for a melanoma-associated cancer antigen, MART-1. Lastly, experimental details will be presented to interface with second generation sequencing platforms for high throughput and quantitative analysis of the reporter barcodes for global cell surface-ome profiling.

1.5 References

1. Stehelin, D.; Varmus, H.E.; Bishop, J.M.; Vogt, P.K. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* **1976**, *260*, 170–173.

2. Comings, D.E. A general theory of carcinogenesis. *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 3324–3328.

3. Melo, J.V.; Barnes, D.J. Chronic Myeloid leukaemia as a model of disease evolution in human cancer. *Nature Rev. Cancer* **2007**, *7*, 441–453.

4. Lugo, T.G.; Pendergast, A.M.; Muller, A.J.; Witte, O.N. Tyrosine kinase activity and transformation potency of bcr-abl oncogene products. *Science* **1990**, *247*, 1079–1082.

5. Daley, G.Q.; Van Etten, R.A.; Baltimore, D. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philidelphia chromosome. *Science* **1990**, *247*, 824–830.

6. Druker, B.J.; et al. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nature Med.* **1996**, *2*, 561–566.

7. Druker, B.J.; et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med.* **2001**, *344*, 1031–1037.

8. Hudis, C.A. Trastuzumab—Mechanism of action and use in clinical practice. *N. Engl. J. Med.* **2007**, *357*, 39-51.

9. Slamon, D.J.; Clark, G.M.; Wong, S.G.; Levin, W.J.; Ullrich, A.; McGuire, W.L. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **1987**, *235*, 177–182.

10. Yun, C.; Mengwasser, K.E.; Toms, A.V.; Woo, M.S.; Greulich, H.; Wong, K.K.; Meyerson, M.; Eck, M.J. The T790M mutation in EGFR kinase causes drug

resistance by increasing the affinity for ATP. *Proc. Natl. Acad. Sci. USA* **2008**, *105(6)*, 2070–2075.

11. Paez, J.G.; et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **2004**, *304*, 1497–1500.

12. Pao, W.; et al. EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 13306–13311.

13. Yang, F. et al. Nf1-dependent tumours require a microenvironment containing $Nf1^{+/-}$ and c-kit-dependent bone marrow. *Cell* **2008**, *135*, 437–448.

14. Massague, J. TGFβ in cancer. *Cell* **2008**, *134*, 215–230.

15. Egeblad, M.; Werb, Z. New functions for the matrix metalloproteinases in cancer progression. *Nature Rev. Cancer* **2002**, *2*, 161–174.

16. Freije, J. M.; et al. Matrix metalloproteinases and tumor progression. *Adv. Exp. Med. Biol.* **2003**, *532*, 91–107.

17. Denko, N.C. Hypoxia, HIF1 and glucose metabolism in the solid tumor. *Nature Rev. Cancer* **2008**, *8*, 705–713.

18. Milosevic, M.; Fyles, A.; Hedley, D.; Hill, R. The human tumor microenvironment: invasive (needle) measurement of oxygen and interstitial fluid pressure. *Semin. Radiat. Oncol.* **2004**, *14*, 249–258.

19. Seruga, B.; Zhang, H.; Bernstein, L.J.; Tannock, I.F. Cytokines and their relationship to the symptoms and outcome of cancer. *Nature Rev. Cancer* **2008**, *8*, 887–899.

20. Brabletz, T.; et al. Variable b-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10356–10361.

21. Franci, C.; et al. Expression of snail protein in tumor—stroma interface. *Oncogene* **2006**, *25*, 5134–5144.

22. Balkwill, F.; Charles, K.A.; Mantovani, A. Smoldering and polarized inflammation in the initiation and promotion of malignant disease. *Cancer Cell* **2005**, *7*, 211–217.

23. Yang, J.; Weinberg, R.A. Epithelial—mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev. Cell* **2008**, *14*, 818–829.

24. Ludwig, J.A.; Weinstein, J.N. Biomarkers in cancer staging, prognosis and treatment selection. *Nature Rev. Cancer* **2005**, *5*, 845–856.

25. Ntzani, E.E.; Ioannidis, J.P. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet* **2003**, *362*, 1439–1444.

26. Pomeroy, S.; et al. Prediction of central nervous system embryonal tumor outcome based on gene expression. *Nature* **2002**, *415*, 436–442.

27. Lossos, I.; Czerwinski, D.; Alizadeh, A.; Wechser, M.; Tibshirani, R.; Botstein, D.; Levy, R. Prediction of survival in diffuse large b-cell lymphoma based on the expression of six genes. *New Engl. J. Med.* **2004**, *350*, 1828–1837.

28. Freije, W.; Castro-Vargas, F.; Fang, Z.; Horvath, S.; Cloughesy, T.; Liau, L.; Mischel, P.; Nelson, S. Gene expression profiling of gliomas strongly predicts survival. *Cancer Res.* **2004**, *64*, 6503–6510.

29. Lamb, J.; Ramaswamy, S.; Ford, H.L.; Contreras, B.; Martinez, R.V.; Kittrell, F.S.; Zahnow, C.A.; Patterson, N.; Golub, T.R.; Ewen, M.E. A mechanism of cyclin D1 actionn encoded in the patterns of gene expression in human cancer. *Cell* **2003**, *114*, 323–334.

30. Chen, G.; Gharib, T.G.; Wang, H.; Huang, C.C.; Kuick, R.; Thomas, D.G.; Shedden, K.A.; Misek, D.E.; Taylor, J.M.G.; Giordano, T.J.; Kardia, S.L.R.; Iannettoni, M.D.; Yee, J.; Hogg, P.J.; Orringer, M.B.; Hanash, S.M.; Beer, D.G.

Protein profiles associated with survival in lung adenocarcinoma. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 13537–13542.

31. Sreekumar, A.; et al. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **2009**, *457*, 910–915.

32. Zhang, Z.; et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res.* **2004**, *64*, 5882–5890.

33. TCGA, Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **2008**, ***455***, 1061–1068.

34. Parsons, D.W. et al. An integrated genomic analysis of human glioblastoma multiforme. *Science* **2008**, *321*, 1807–1812.

35. Hanash, S. Disease proteomics. *Nature* **2003**, *422*, 226–232.

36. Hanash, S. Integrated global profiling of cancer. *Nature Rev. Cancer* **2004**, *4*, 638–643.

37. Balic, M.; Dandachi, N.; Hofmann, G.; Samonigg, H.; Loibner, H.; Obwaller, A.; van der Kooi, A.; Tibbe Arjan, G.J.; Doye Derald, V.; Terstappen Leon, W.M.M.; Bauernhofer, T. *Cytometry, Part B* **2005**, *68*, 25–30.

38. Adams, A.A.; Okagbare, P.I.; Feng, J.; Hupert, M.L.; Patterson, D.; Gottert, J.; McCarley, R.L.; Nikitopoulos, D.; Murphy, M.C.; Soper, S.A. Highly efficient circulating tumor cell isolation from whole blood and label-free enumeration using polymer-based microfluidics with an integrated conductivity sensor. *J. Am. Chem. Soc.* **2008**, *130(27)*, 8633–8641.

39. Nagrath, S.; et al. Isolation of rare circulating tumour cells in cancer patients by microchip technology. *Nature* **2007**, *450*, 1235–1239.

40. Maheswaran, S.; et al. Detection of mutations in EGFR in circulating lung-cancer cells. *N. Engl. J. Med.* **2008**, *359*, 366–377.

41. Kallenbach, N.R.; Ma, R.I.; Seeman, N.C. An immobile nucleic acid junction constructed from oligonucleotides. *Nature* **1983**, *305*, 829–831.

42. Rothemund, P.W.K. Folding DNA to create nanoscale shapes and patterns. *Nature* **2006**, *440*, 297–302.

43. Winfree, E.; Liu, F.; Wenzler, L.A.; Seeman, N.C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **1998**, *394*, 539–544.

44. Benenson, Y.; Gil, B.; Ben-Dor, U.; Adar, R.; Shapiro, E. An autonomous molecular computer for logic control of gene expression. *Nature* **2004**, *429*, 423–429.

45. Shin, J.S.; Pierce, N.A. A synthetic DNA walker for molecular transport. *J. Am. Chem. Soc.* **2003**, *125*, 10834–10835.

46. Robinson, B.H.; Seeman, N.C. The design of a biochip—a self-assembling molecular-scale memory device. *Protein Eng.* **1987**, *1*, 295–300.

47. Sano, T.; Smith, C.L.; Cantor, C.R. Immuno-PCR: very sensitive antigen detection by means of specific antibody-DNA conjugates. *Science* **1992**, *258*, 120–122.

48. Zhou, H.; et al. Two-color, rolling-circle amplification on antibody microarrays for sensitive, multiplexed serum-protein measurements. *Genome Biol.* **2004**, *5*, R28.

49. Fredriksson, S.; et al. Multiplexed protein detection by proximity ligation for cancer biomarker validation. *Nat. Methods* **2007**, *4*, 327–329.

50. Nam, J.M.; Thaxton, C.S.; Mirkin, C.A. Nanoparticle-based bio-bar codes for the ultrasensitive detection of proteins. *Science* **2003**, *301*, 1884–1886.

51. Kattah, M.G.; Coller, J.; Cheung, R.K.; Oshidary, N.; Utz, P.J. HIT: a versatile proteomics platform for multianalyte phenotyping of cytokines, intracellular proteins and surface molecules. *Nature Med.* **2008**, *14*, 1284–1289.

52. Melkko, S.; Scheuermann, J.; Dumelin, C.E.; Neri, D. Encoded self-assembling chemical libraries, *Nat. Biotech.* **2004**, *5*, 568–574.

53. Blattman, J.; Greenberg, P. Cancer immunotherapy: a treatment for the masses. *Science* **2004**, *305*, 200–205.

54. Schumacher, T.N. T-Cell-receptor gene therapy. *Nat. Rev. Immunol.* **2002**, *2*, 512–519.

55. Morgan, R.A.; Dudley, M.E.; Wunderlich, J.R.; Hughes, M.S.; Yang, J.C.; Sherry, R.M.; Royal, R.E.; Topalian, S.L.; Kammula, U.S.; Restifo, N.P.; et al. Cancer Regression in Patients After Transfer of Genetically Engineered Lymphocytes. *Science* **2006**, *314*, 126–129.