Published in final edited form as:

J Proteomics. 2014 July 31; 0: 103–112. doi:10.1016/j.jprot.2014.04.012.

A New Class of Protein Cancer Biomarker Candidates: Differentially-Expressed Splice Variants of ERBB2 (HER2/neu) and ERBB1 (EGFR) in Breast Cancer Cell Lines

Gilbert S. Omenn, M.D., Ph.D., Yuanfang Guan, Ph.D., and Rajasree Menon, Ph.D.

University of Michigan Center for Computational Medicine and Bioinformatics

Abstract

Combined RNA-Seq and proteomics analyses reveal striking differential expression of splice isoforms of key proteins in important cancer pathways and networks. Even between primary tumor cell lines from histologically-similar inflammatory breast cancers, we find striking differences in hormone receptor-negative cell lines that are ERBB2 (HER2/neu)-amplified versus ERBB1 (EGFR) over-expressed with low ERBB2 activity. We have related these findings to protein-protein interaction networks, signaling and metabolic pathways, and methods for predicting functional variants among multiple alternative isoforms. Understanding the upstream ligands and regulators and the downstream pathways and interaction networks for ERBB receptors is certain to be important for explanation and prediction of the variable levels of expression and therapeutic responses of ERBB+ tumors in the breast and in other organ sites.

Alternative splicing is a remarkable evolutionary development that increases protein diversity from multi-exonic genes without requiring expansion of the genome. It is no longer sufficient to report up- or down-expression of genes and proteins without dissecting the complexity due to alternative splicing.

Keywords

breast cancer subtypes; ERBB2 (HER2/neu); ERBB1 (EGFR); splice variant transcripts; splice variant proteins; pathway analyses

1.0 Introduction

There is now overwhelming evidence that nearly all multi-exonic genes in multi-cellular organisms produce a mixture of transcript products through the process of alternative splicing at the premRNA (heterogeneous nuclear RNA) stage. These splice transcripts may

© 2014 Elsevier B.V. All rights reserved.

Correspondence should be addressed to Dr. Gilbert S. Omenn, Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109-2218, USA.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

be translated into splice variant proteins. The evolution of splicing facilitated much greater protein diversity without an increase in the extent of the genome coding for proteins.

In a wide range of cancers, distinctively misregulated splicing contributes to the neoplastic behavior of the cells. Zhang and Manley reviewed striking examples of exon-skipping in *RON*, multiple exon skipping in *BRAF*, exon inclusion in *SYK*, alternative 5' splice sites in exon 2 of *BCL-X_L* vs *BCL-X_S* dimers, alternative 3' splice sites in *VEGF*, intron retention in *STAT2*, mutually exclusive exons 9 and 10 in pyruvate kinase M2, and 40 splice variants of *MDM2* affecting binding to p53[1].

In this paper prepared for the 10th Siena Meeting “From Genome to Proteome: 20 Years of Proteomics”, we review the emergence of splice variant proteins as a new class of cancer biomarker candidates, with a focus on breast cancers induced by amplification of *ERBB2* at chromosome 17q12. Our work on chromosome 17 is a part of the Chromosome-centric Human Proteome Project (C-HPP)[1, 2] of the HUPO Human Proteome Project (HPP, www.thehpp.org)[3-5].

2.0 ERBB2 (HER2/neu) and ERBB1 (EGFR)

The epidermal growth factor receptor family, one of 20 families of human receptor tyrosine kinases (RTK), is often over-expressed in human cancers. These 185 kDa transmembrane glycoproteins are activated by EGF, EGF-like, or neuregulin ligands to form homodimers and heterodimers, and the intracellular tyrosine kinase domain then activates several important oncogenes downstream in the signaling cascade. ERBB2 is unique in lacking a ligand; its signaling depends upon heterodimer formation. Several clinically effective small molecule and protein drugs target the EGFR (ERBB1) or HER2/neu (ERBB2) receptors. As part of the Chromosome 17 team of the Chromosome-centric Human Proteome Project (C-HPP)[2, 6-8], we have utilized proteomics and RNA-sequencing to investigate and annotate evidence of alternative splicing and downstream pathways activation[9-11]. After first studying the mouse model of ERBB2-induced breast cancers[12], we turned to the human ER-/PR- breast cancer cell lines SKBR3, SUM149, and SUM190. These cell lines express transcripts for ERBB2 at 300, 14, and 400 reads per kilobase per million mapped reads (RPKM) and for ERBB1 at 1, 60, and 0.6 RPKM[10], respectively. Thus, SKBR3 and SUM190 are highly expressing ERBB2, and SUM149 is highly expressing ERBB1 (Table 1).

2.1 The ERBB2 (HER2/neu)-Induced Breast Cancer Subtype

About 15-20 percent of human breast cancers are due to amplification and/or high expression of the *ERBB2* (*HER2/neu*) gene at Chromosome 17q12. This gene encodes a cell surface receptor whose activation drives downstream signaling and cell proliferation. The monoclonal antibody drug trastuzumab (Herceptin) targets this receptor and is effective in treating patients with this subtype of breast cancers.

Chromosome 17 has the second highest density of genes among the chromosomes (after Chr 19). It contains many regions with multiple cancer-associated genes (based on GeneCards), primarily on the long arm (17q), as documented in Table 1 of Liu et al, 2013[9]. These

genes include the ubiquitous tumor suppressor *TP53* and the oncogenes *BRCA 1*, *NF1*, and *ERBB2*. In all, there are 44 cancer-associated genes, of which 36 occur in regions with gene density >30 genes/Mb and all oncogenes had >5 other cancer-associated genes in proximity.

Using the Oncomine resource for transcript datasets from numerous cancers, we demonstrated that 13 of the top 20 expressed genes associated with *ERBB2*-induced breast cancers are on Chromosome 17 (which has 4.5% of the protein-coding genes), nearly all quite near *ERBB2* (*HER2/neu*) at 17q12. They form a variable amplicon, with 2 to 23 genes amplified and co-expressed in various tumors. When we compared the top 20 expressed genes on chromosome 17 for *ERBB2*+, *ER*+/PR+, and triple-negative breast cancers, we found no overlap at all among these three sets of 20 genes[9]; the different breast cancer cell types are quite different. The *ERBB2* amplicon typically occurs over a 1.5Mb region 17q11.2-q21.2; it may be observed as homogeneously staining regions or extrachromosomally as double minute chromosomes or submicroscopic episomes[13].

Diagnosis and treatment of *ERBB2*-induced cancers is not limited to breast cancers; in fact, we increasingly recognize that diagnosis and treatment should be based on the mechanism involved, not the site of origin of the tumor mass. Significant percentages of gastric cancers and colorectal cancers, plus rare lung cancers, are *ERBB2*+; they often respond well to the drug trastuzumab, even though there are tissue-specific modifying factors that may reduce the tumor response in patients.

2.2 The Chodosh Mouse Model of *ERBB2*-induced breast cancers

We reported[12] 540 known and 68 novel splice variants from LC-MS/MS datasets of mouse *ERBB2*+ mammary tumor and normal mammary tissue[14]. These variants reflected multiple mechanisms, including new translation start sites, new splice sites, extension or shortening of exons, deletion or switch of exons, intron retention, and translation in an alternative reading frame[12]. For a subset of 32 of the 45 novel splice variants detected only in tumor tissue, qRT-PCR was performed and confirmed presence of the corresponding mRNA for 31, with agreement on the expression difference at the protein level in 29.

Many interesting splice variants were annotated, including variants of two different proteins that interact with the breast cancer-predisposing gene *BRCA1*: A splice peptide sequence from the second intronic region of the leucine-zipper-containing LZF (rogdi) and a splice peptide sequence from the first intronic region of transcription factor SOX7. Both were shown to have the *LIG_BRCT_BrcA1_1* phosphopeptide motif that directly interacts with the carboxy-terminal domain (BRCT) of *BRCA1*.

Of 15 biomarker candidates over-expressed in the tumor tissue lysate as reported in the original study[14], we were able to identify 10 among our proteins with splice variants; further studies are needed to deduce and confirm their functional consequences. We performed a protein interaction network analysis and visualization with Cytoscape and Michigan Molecular Interactions (MiMI), displaying direct protein-protein interactions for 179 proteins from the tumor sample. Striking were interactions between differentially-expressed variants of cell division cycle 42 (*CDC42*), radixin (*RDX*), arhgdia, and methionyl amino peptidase (*METAP2*). Arhgdia interacts with the ezrin/rdx/moesin (*ERM*)-

CD44 system to initiate activation of Rho family members, including CDC42, followed by actin filament reorganization and increased cell motility.

In our mouse model study, we used a peptide-to-protein algorithm to create the list of proteins (< 1% FDR) for the peptides identified [12, 15]. The algorithm produced a minimally redundant list of 540 proteins known in Ensembl based on proteotypic peptides, protein length, and total number of matching peptides. Proteotypic peptides were identified for 110 out of 540 known Ensembl proteins identified from the tumor samples. Due to the inherent bias toward peptides of higher concentrations, most peptides identified from a mass spectrometric analysis are shared by multiple variants of a gene, as well as multiple homologous proteins. We found homologous human proteins in the ERBB2+ SKBR3 and SUM 190 human cell line data for 49 and 38 of these 110 proteins, respectively. The homologous variants, ANXA6_673 (ENSP00000346550), BLVRB_206 (ENSP00000263368), IDH2_452 (ENSP00000331897) and OGDH_1063 (ENSP00000222673) were found in common in SKBR3 and SUM190 (the number following the gene symbol is the variant protein length). BLVRB and OGDH are involved in the TCA cycle suggesting a common energy process in these two cell lines.

2.3 ERBB2+ Human Cancer Cell Lines

Alternatively-spliced transcripts (ASTs) and proteins (ASPs) in six ERBB2+ cancer cell lines: colorectal (LIM2405, LIM1899), gastric (KATOIII, SNU16), and breast (SUM149, SUM190) were studied as part of our initial Chr 17 C-HPP paper [9] for the Chromosome-centric Human Proteome Project [8]. In total, Liu et al reported 195 distinct ASTs from 144 genes. As shown in Figure 1a, SUM190 had remarkably high expression of the shorter variant of ERBB2 (translated product of 1225 aa length), with much lower, but detectable reads (RPKM 1-2) in four of the other five cell lines. Figure 1b shows highly differential transcript expression of the longest splice variant of each of three genes (CDK12, FBXL20, GRB7) that are part of the *ERBB2* amplicon at 17q12.

A total of 30 distinct splice variants of 11 genes that are located in the 17q12 region were identified from one or more of the three human breast cancer cell lines SKBR3, SUM190 and SUM149 by Menon et al [11]: genes (5' to 3') including *AP2B1*, *TAF15*, *ACACA*, *PSMB3*, *RPL23*, *LASPI*, *RPL19*, *ERBB2*, *MIEN1*, *GRB7* and *ORMDL3*. The criteria for splice variant identification used in these analyses, in contrast to the mouse model, were the presence of unique reads aligning specifically to the transcript and a matching peptide, thus integrating transcriptomics and proteomics. The heat map of the relative expressions of these transcripts shows SUM190 and SKBR3 as more similar to each other than to SUM149 (Figure 2); this similarity is mainly due to the overexpression of multiple variants of ERBB2 itself, not, so far as we can tell, from a pattern of overexpression across many of these nearby genes [11].

Unlike the findings in Liu et al using just RNA-seq data, no variants of CDK12 and FBXL20 were identified in these Menon et al analyses as the requirements for peptides matching to the translated products of these genes were not met. This could be due to low expression of the protein products of these genes, which would be consistent with the relatively low transcript expression in Liu et al [9] (Figure 1b).

Variant MIEN1_115 (ENSP00000377778) was found with similar expression levels in SKBR3 and SUM190 compared to its lower expression in SUM149; *MIEN1* is situated between *ERBB2* and *GRB7* in the 17q12 amplicon region. Katz et al reported the oncogenic potential of MIEN1 in invasive breast cancers; high levels of MIEN1 protein were associated with colony growth in soft agar, invasion into collagen matrix and formation of large acinar structures[16]. On the other hand, four variants of *ORMDL3*, another gene in 17q12, were expressed in SKBR3, but not in SUM190 or SUM149 (Figure 2); functional relevance of *ORMDL3* in breast cancers has not been reported.

Another interesting observation was the expression of two variants of *LASP1* found in SUM149 and their absence in the *ERBB2* over-expressed cell lines, SKBR3 or SUM190. This observation is supported by a study finding *LASP1* highly over-expressed in invasive carcinomas compared to fibroadenomas; in contrast to previous reports, there was no relation between *LASP-1* protein level and HER-2/neu (*ERBB2*) protein expression[17].

There are also multiple families of related genes on chromosome 17 occurring in clusters and reflecting tandem gene duplication, e.g., growth hormone, Schlafen growth regulators, olfactory receptors, cytokines, chemokine ligands, keratins, keratin-associated proteins, homeobox and chromobox proteins [9]. Band q21.2 contains 109 genes, of which 50% represent keratins (28) or keratin-associated proteins (25).

2.4 Pathways Analysis and Protein-Protein Interactions in SUM149 and SUM190 Inflammatory Breast Primary Cell Lines

The same three breast cancer cell lines were deployed in this study which emphasized pathways specific for *ERBB1* in SUM149 and for *ERBB2* in SUM190 [10]. RNA-Seq was used to identify a total of 31 oncogenes with significant transcript levels in these three cell lines. It is now feasible to quantitate >10,000 transcripts and proteins in cell line studies [18] [19], but 1000-3000 proteins are much more typical of shotgun proteomics analyses, as in our own studies. However, the proteome findings are important in documenting that the transcripts were actually translated, in permitting detection of post-translational modifications, and in finding the occasional protein in the absence of measurable mRNA. Zhang et al[10] identified 1444, 1396, and 964 proteins (with 2 or more peptides and FDR<1%), respectively, in SKBR3, SUM149, and SUM190 cell lysates. The number of observed interactors for each oncogene ranged from 4% for *JAK1* to 27% for *MYC*; the percentage reflects the total protein interactions observed experimentally divided by the number in STRING (Search Tool for the Retrieval of Interacting Genes/Proteins, v9.0) and I2D (Interologous Interaction Database, v1.95). The key signaling pathways were *ERBB* signaling, *EGFR1* signaling, integrin outside-in signaling, and validated targeted of c-MYC transcriptional activation.

SUM190 has high transcript levels of *ERBB2* and *ERBB3*, low levels of *ERBB4* and undetectable *ERBB1* (*EGFR*), with a high value for ligand *HBEGF* and low value for the *EGFR* ligand amphiregulin (*AR*). The high *ERBB3* may be quite important in cell proliferation via this pathway [20]. The downstream pathways activated include *Crk/Abl*, *FAK*, *PAK/Jun*, *Slc/Grb2/Fas/Raf/MEK/ERK/MYC/JNK*, and *PI3K/PKB/Akt/mTOR*, p27 [10, Figure 2]. In contrast, SUM149 has high *EGFR* (*ERBB1*), very low *ERBB2*,

moderately high ERBB3, and low ERBB4 transcript reads, with high expression of its ligands amphiregulin, epiregulin (EPR), and transforming growth factor alpha (TGFA). Its downstream pathways activated are PLCgamma, CAMK, STAT3, Src/FAK, Crk/Abl, Nck/Pak/JNKK/JNK/Jun/Elk, Slc/Grb2, Ras/Raf/MEK/ERK/MYC/Elk, STAT5, PI3K, PKB, mTOR, p27, and p21.

Thus, Zhang et al presented a composite of transcriptomic, proteomic, and interaction data for SUM149 and SUM190 [10]. As shown in Figure 3, MYC stands out as having high transcript reads without any proteomic signal in either cell line, yet many protein interactors. Short lines outward from ERBB2 in the center indicate strong interactions for oncogenes EGFR, ERBB3, ERBB2IP, GRB2, GRB7, and KRAS; level of transcript reads is proportional to the size of the circles, while a black perimeter on the circle indicates proteomic evidence. MYC, GRB2, and EGFR have the highest numbers of interactions with observed proteins.

2.5 Pathway Enrichment Studies using Gene Ontology and BioCarta

Menon et al [11] extended the preceding studies of SKBR3, SUM149, and SUM190 cell lines. We identified multiple splice variants (4406 transcripts) from a total of 1167 distinct genes, including ERBB2 and EGFR. Many variants are expressed at quite different levels in the three breast cancer cell lines. None of the six splice variants of EGFR was found in SUM190, whereas SUM149 expressed all six and SKBR3 expressed five, including the long canonical variant EGFR-1210. Menon et al [11] extensively characterized the EGFR variants and used I-TASSER folding algorithms to interpret the loss of exon 4 and exon 28 in the highly expressed EGFR-1091 compared with EGFR-1210. The sum of the shorter variants exceeds the number of reads of the highly expressed EGFR-1210 in SUM149. A heat map of BioCarta Geneset enrichment shows clustering of the ERBB2-over-expressing SKBR3 and SUM190; multiple variants of ERBB2 and of STAT3 were expressed in SKBR3 and SUM190, but not in SUM149. As noted by Zhang et al [10] downstream ERBB receptor signaling notably affects cell cycle, cell adhesion, cell motility, and apoptosis.

Heat maps indicate that the inflammatory breast cancer cell lines SUM190 and SUM149 are similar in mRNA splicing; across the three cell lines we found multiple variants of splice factors *eftud2*, *nbp211*, *pbp2*, *ptbp1*, *snpra*, *snrpd2*, *snrpe*, and *ybx1*. Much additional work is needed to place such variants into the very complex spliceosome machinery and well-defined steps in splicing out introns and ligating exons. Since these two cell lines are markedly different in ERBB2, the congruence for splicing variants must reflect some other pathway than ERBB2-mediated downstream effects.

Shared enrichment between the inflammatory cancer cell lines SUM149 and SUM190 included vesicle ATP binding, GTP binding, RNA binding, citric acid cycle, aminoacyl-tRNA synthesis, RNA translation, protein localization and Ras GTPase activity. By contrast, the SKBR3 and SUM190 high ERBB2+ cell lines had 138 transcripts in common from 92 genes identified with multiple unique reads that were not in SUM149; DAVID indicated enrichment for electron transport chain, intracellular transport, and phosphate metabolism. Vesicular protein trafficking may involve Annexin 6, whose splice variants we have studied extensively in the Chodosh mouse model [12] and with I-TASSER computational methods

[21]. We also found many cell line-specific GO biological process terms and BioCarta pathways: amino-sugar metabolism, caspase activity, arrestin activation of MAP kinases, and endocytosis by NDK, phosphins and dynamins in SKBR3; lipid metabolism in SUM190; and cell adhesion, integrin signaling, Erk1/Erk2/Mapk signaling, K48-linked ubiquitination, and translational control by eIF4e and p70S6 in SUM149. The total distinct read counts of FASN in SUM190 and SKBR3 were higher compared to SUM149 (8055 and 5846 versus 4400). The distinctively high fatty acid synthase (FASN) read count in SUM190 suggests a more prominent role of lipid metabolism in SUM190 homeostasis. It has been shown that ERBB2 overexpression increases translation of FASN [22]. Kumar-Sinha et al [23] reported that HER2 stimulates the FAS promoter and ultimately mediates increased fatty acid synthesis through a phosphatidylinositol 3'-kinase-dependent pathway.

All three cell lines had networks enriched with protein interactions that had ubiquitin C at the center; at least one unique splice variant of UBC was found in each breast cancer cell line, whereas all three had the canonical long variant (UBC_685) with its 9 ubiquitin domains (Figure 4). In contrast, the unique variants specific to SKBR3 and SUM190 have two Ub domains and the unique variant in SUM149 has eight.

Rho signaling plays a key role in tumor cell motility and in inflammatory breast cancers, especially Rho C and Rho A [24]. We noted that phosphorylated CAV1 is an effector of Rho/ROCK signaling to promote late-stage tumor progression and metastasis; we identified multiple variants of CAV1, ITGB1, and SRC in SUM149 only [11]. Merajver and colleagues have begun to distinguish the roles of highly homologous RhoC and RhoA GTPases and the different roles of RhoC in different cell lines (MDA231 vs SUM149)[25].

3.0 Predicting Differential Functions of Splice Isoforms from RNA-seq Data

Guan and colleagues have developed algorithms to identify the “responsible isoform(s)” of a gene and classify models at the isoform level instead of only at the gene level. Our mouse ERBB2-induced mammary cancer datasets, plus protein structure modeling and experimental evidence, have been used to validate some of these predictions [26]. This method is extendable to additional supervised learning algorithms for large-scale genomic data integration using support vector machines (SVM), Bayesian classification, or artificial neural networks. There are often opposing or quite different functions for pairs of splice isoforms, such as selectivity for different cations in TRPM3 channels, pro- vs anti-apoptotic actions for BCLX and CASP3 isoforms, and activation versus repression of transcription with OSR2 isoforms [27-29]. Domain analysis can be quite useful, as protein evolution often appears to be focused on domains more than whole protein sequences. As of the end of 2012, however, Eksi et al found that only 34% of the isoform pairs in the NCBI database have different domains; more subtle differences must be used to predict functional differences [26].

Various methods utilize binding regions and specific binding sites. Our approach iteratively corrects the composition of a set of isoforms of annotated genes for a specific function through matching to “positive” and “negative” isoforms and maximizing the difference between the positive and negative sets [26]. This approach is termed a “multiple instance

learning task”, aimed at identifying the “hidden labels” of the isoforms of the positively annotated genes and then using these labels to construct classification models for additional isoforms. Numerous RNA-seq datasets were examined for co-expression patterns and potential co-functionality as organized in Gene Ontology with 1792 biological process terms and 20-300 genes annotated to each. For some biological processes, interpreting data at the isoform level with MIL significantly improved the prediction performance compared with gene level SVM results.

Although it has been estimated that 95% of multi-exon genes have splice isoforms [30], only 13% of the genes are documented with validated multiple isoforms in NCBI. Results were better for known multi-isoform genes than for those reported to have only one isoform, and results were robust across a wide range of gene expression levels and GO term size. An important principle is that the functional isoform(s) of a gene must be expressed at the protein level in normal physiological conditions in order for the biological functions of the transcripts to be active in the cell; in normal breast tissue, proteomics results support this principle. The power of the method was tested further with examples in which isoforms of a single gene are assigned unrelated biological functions. Two CDKN2a splice isoforms have different reading frames; NM_001040654.1 (168 aa) enhances p53-dependent transactivation and is pro-apoptotic, while NM_009877.2 (169aa) regulates the transmembrane receptor protein serine/threonine kinase signaling pathway; the first has five ankyrin repeats, while the latter has a cyclin CDK4-dependent kinase inhibitor N-terminus domain and no ankyrin domains. I-TASSER predicts drastically different 3D structures. At a higher level, both isoforms shared the function of regulation of G1/S transition in the mitotic cell cycle, as predicted by the algorithm.

A contrasting example is annexin 6 (ANXA6); NM_013472.4 and NM_001110211.1 have very similar structures, as we have reported in great detail [4, 21]. A difference of six amino acids (525-530) affects the accessibility to Thr-535 and Ser-537, making NM_013472.4 more likely to undergo phosphorylation, which is predicted independently by our function prediction algorithm. Note that such predictions are based on the RNA-seq data alone. Treating the “gene expression” levels without specific analysis of the levels and functions of the individual splice transcripts and splice proteins would be unreliable and would ignore data available from RNA-seq analyses.

Multiple variants of STAT3 were expressed in the ERBB2+ cell lines SKBR3 and SUM190: five in SKBR3 (STAT3_770, STAT3_769, STAT3_722, STAT3_672, STAT3_84) and three long isoforms in SUM190 (STAT3_770, STAT3_769, STAT3_722) [11] (Figure 5). STAT 3 was not detected in SUM149. STAT3 is located in the 17.q21 band and annotated as a transcription factor. Since STAT3_770, STAT3_769 and STAT3_722 variants have mouse homologous proteins, we used the MIL algorithm to predict the potential functions of these three variants. According to the function predictions, the top-ranking GO terms for these variants were distinct, with remarkably high fold enrichment. Ventricular cardiac muscle tissue development (GO:0003229) was the top ranking term with 31-fold enrichment for the STAT3_770. Interestingly, regulation of the EGFR signaling pathway (GO:0042058) was among the top ranking processes (11-fold) compared to only slight enrichment for STAT3_769 and STAT3_722. This observation suggests that the specific role of

STAT3_770 is in ERBB signaling. Lipid homeostasis (GO:0055088) was the top ranking GO term (28- fold) for STAT3_769, while STAT3_770 had 2-fold and STAT3_722 had 1.4-fold enrichment, respectively. Fatty acid beta-oxidation (GO:0006635) was the top ranking term for STAT3_722 with 49-fold enrichment compared to 4-fold in STAT3_770 and 6 fold in STAT3_769. In a recent study [31], STAT3 activation by phosphorylation was observed in Her2 over-expressing human breast cancer cell lines. Furthermore, the STAT3 inhibitor Stattic abolished the cancer stem cell phenotype in HER2+ breast cancers[31].

These algorithms can be applied to many aspects of our recent findings. For example, in the mouse HER2/neu-amplified mammary tumor model, Whiteaker et al [14] had confirmed 15 biomarker candidates from shotgun mass spectrometry as over-expressed in tumor lysates using MRM targeted mass spectrometry. We then found that 10 of the 15 had splice variants; it will be interesting to evaluate the functional features of these different isoforms. For example, the soluble long isoform epidermal growth factor receptor ERBB1-1210 is the source of the major circulating epidermal growth receptor in human blood [32]. As predicted by our MIL algorithm [26], ERBB1-1210 is 10 times more likely to be involved in regulation of coagulation (GO:0050818), than the shorter variant ERBB1-655. Moreover, ERBB1_1210 is predicted to be involved in cholesterol metabolic process (GO:0008203) and response to steroid hormone stimulus (GO:0048545), while ERBB1_655 is not. The sEGFR has been associated with stage I/II and stage III/IV epithelial ovarian cancer, where its concentration in patients is significantly lower than in healthy women [32] and may complicate the efficacy of EGFR-targeted therapies [33].

We are currently extending the MIL algorithm to human applications, which will facilitate further such predictive analyses.

4.0 Nuclear Topology of Chromosomally-Active Segments

We are exploring the nuclear topological features of these contrasting patterns of expression of ERBB2 and ERBB1 involving their chromosomal segments at 17q12 and 7p11.2, respectively [Rajapakse I & Omenn GS, unpublished data].

A longterm goal is to understand the multiple levels of regulation of gene expression, including transcription factors and their binding site patterns; non-protein-coding DNA elements involved in nuclear organization, transcriptional regulation, and cellular differentiation, as investigated by the Encyclopedia of DNA Elements (ENCODE) initiative; epigenetic modifications of DNA and histones; and, finally, alternative splicing. Using sequence probes, fluorescent labels, cross-linking reagents, and confocal microscopy, researchers have now mapped the 3D or 4D topology of the nucleus. The architecture of the interphase nucleus is non-random. Heterochromatin is enriched at the nuclear lamina, transcription and splicing aggregates are organized in foci, and distinct chromosome territories occupy specific 3D spaces. Moreover, the topology is dynamic, creating a 4D spatiotemporal nucleome. Gene-rich chromosomes and transcriptionally active chromosomal segments (euchromatin) are more central, whereas gene-poor chromosomes and transcriptionally inactive segments (heterochromatin) are more peripheral. This

distribution is conserved in evolution. Effects of epigenetic modifications on the 3D architecture of the nucleus have not yet been characterized.

We predict that chromosomal segments at 17q12 and 7p11 would be in juxtaposition and central if both ERBB2 and ERBB1 were highly expressed; the phenomenon could be magnified by amplicon-based co-expression of the genes associated with ERBB2. However, the extensive aneuploidy of the cancer cell lines may make the 3D nucleome observations particularly challenging. Maybe the SUM149 and SUM190 cell lines from primary tumors are less aneuploid than SKBR3 and other established cell lines. Our current activity is to begin to characterize these ERBB2/ERBB1 interactions with specific probes and fluorescent labels in primary fibroblasts (IMR90), the human mammary epithelial cells (HMEC), and other common cell lines, such as erythroleukemia cell line K562 and lymphoblastoid cell line GM066990, before moving to the cell lines we have characterized with RNA-seq and proteomics.

Thomas Ried and colleagues at the National Cancer Institute have proposed an initiative to generate comprehensive 3D maps of the interphase nucleus of cells in distinct functional states and with specific mutations and epigenetic modifications using chromosome conformation capture, interphase FISH, and high-resolution microscopy. They will compare such maps to maps from cells at other physiological and pathological states, explore the functional role of chromatin remodeling, uncover the mechanisms governing the establishment of lineage specific 3D conformation in normal cells and how these are perturbed by disease, develop bioinformatic tools to interrogate these data in the context of multicellular organisms, cellular differentiation and disease, and develop a publicly accessible reference database[34,35]. Clearly, compartmentalization of biological processes, accessibility of chromatin, and spatial sequestration of genes and their regulatory factors modulate the output and functional status of genomes. The Misteli lab [36] has created a spatiotemporal framework for detecting the formation of translations in living cells within hours of occurrence of double-strand breaks; these translocations form preferentially between prepositioned genome elements as key factors of the DNA repair machinery uncouple DSB pairing from translation formation. Our overall aim is to reach beyond a linear interpretation of genome sequence to dimensions including time and space in regulation of gene expression and translation to proteins.

5.0 Discussion

5.1 Misregulation of alternative splicing in cancers by regulatory splicing factors

Genome-wide studies indicate at least 860 RNA-binding proteins (RBPs), of which only two dozen or so are well-characterized [1]. The regulation of splicing consists of selecting splice sites in premRNA transcripts through cis-acting regulatory sequences and trans-acting RBPs. Two families are well-studied: serine/arginine-rich SR proteins, which bind to exonic (ESE) and intronic (ISE) splicing enhancers and promote exon inclusion; and heterogeneous nuclear ribonucleoproteins hnRNPs, which bind to exonic (ESS) and intronic (ISS) splicing silencers and lead to exon skipping. For example, SRSF1 is up-regulated in many human tumors and mammary epithelial cells; it acts as a proto-oncogene through many target pre-mRNAs, including deltaRON (expressing exon 12, skipping exon 11), BIN1, S6K1, and

mTOR[1]. SRSF3 also is over-expressed in many tumors, and regulates splicing of p53. hnRNP A1 and A2 each potentially regulates >2000 splicing events, including one-third in common, such as PKM together with PTB; they bind to sequences flanking exon 9 of pyruvate kinase pre-mRNA, repress exon 9 inclusion, and promote exon 10 inclusion instead (producing the PKM2 associated with the Warburg effect of aerobic glycolysis, and proliferation of the tumor). Knockdown of A1/A2 and PTB, or of MYC, drives PKM2 to PKM1. PTB also regulates splicing of USP5, a deubiquitinating enzyme whose knockdown leads to accumulation of p53. In many cases, levels and activities of RBP2 changed substantially without any change in the overall levels of the set of transcripts from particular genes. Only isoform analysis will detect these important changes that can be drivers of cancers.

Splicing factors themselves can be mutated or be subjected to alternative splicing. Among the most frequently mutated genes in myelodysplastic syndromes [37] are SF3B1, U2AF1, SRSF2, and ZRSR2, which all are involved in the selection of splice sites at the 3' end of introns, leading to intron retention; if the intron has stop signals, nonsense-mediated mRNA decay may be activated, making these sequences undetectable.

Splicing may be quite different between mouse and human, as reported for phenotypes associated with SF3B1 and SRSF2 (both 100% identical between mouse and human), U2AF1 (96% identical), and ZRSR2 (82% identical); only one quarter of human AS events were observable in the mouse [37, 38]. The actions of a spliceosomal mutation may be opposite in different cellular contexts, so predictions are quite challenging. There are drugs that are modulators of spliceosome functions (sudemycins, pladienolide B, FR901464, spliceostatin A). Binding of FR901464 to SF3b promotes retention of intron 1 of p27, a cyclin-dependent kinase inhibitor, producing a C-terminal truncated protein isoform resistant to proteasomal degradation, which inhibits CDK2 and limits cell growth[39]. Sorting out variability due to cellular heterogeneity, different neoplasms, and different host cells will be very challenging.

More than one isoform for each of hnRNPA1, PTBP1, SRSF1 and SRSF3 were found in SKBR3, SUM190 and SUM149 [11]. Gene Ontology terms enriched for the splice variants identified in these three breast cancer cell lines included splicing. The multiple variants of snRNPs and other splice factors identified may play distinct roles in the splicing patterns of each of these cell lines. The transcript profiles for mRNA splicing show SUM190 and SUM149 profiles clustered together (supplementary document part 4 in[11]); such similar splicing mechanisms may not be due to the effect of ERBB2 downstream signaling, since SUM149 does not over-express ERBB2.

5.2 Evolutionary aspects of alternative splicing

Tissue-specific and lineage-specific splicing appear to have accelerated evolution, as proposed by Walter Gilbert in 1978, and are much more conserved than species-specific differences in splicing patterns. As summarized by Papasaikas and Valcárcel [41], different cell types can interpret the same sequence of a pre-mRNA as either an exon or an intron. This leads to cell type-specific alternative interpretations of the genomic information. Alternative splicing allows the shuffling of protein-coding domains and confers distinct

sensitivity of the spliced mRNAs to regulatory factors. Thus, gene transcription and alternative splicing provide complementary mechanisms by which particular cell types can determine the complement of proteins required for carrying out their specialized functions in the organism, with far more transcription factors than major splicing regulators. Barbosa-Morais *et al.* [37] and Merkin *et al.* [38, 39], found that alternative splicing patterns are dominated by species-specific differences that accumulate even during relatively short evolutionary periods of 6 million years, implying that tissue-specific splicing has diverged in particular lineages at a pace one to two orders of magnitude faster than transcriptional changes. Nevertheless, both groups uncovered a few hundred alternatively spliced exons whose tissue-specific regulation is highly conserved, in some cases over periods of hundreds of millions of years.

Merkin *et al.* [39] discovered enrichment in binding sites for well-known regulators of alternative splicing during cell differentiation, uncovering an ancestral splicing code. By inserting a human chromosome 21 into mouse cells, they were able to show that the human pattern of splicing was retained, presumably driven by the species-specific sequences in the target genes. In addition, evolution of splicing patterns accelerated within primates. Species-specific splicing is more frequent in regulatory genes encoding nucleic acid binding proteins; both tissue- and species-specific exons typically occur in unstructured regions of proteins involved in protein-protein interactions.

6.0 Concluding Remark

Alternative splicing is a remarkable evolutionary development, only recently appreciated. Given the role of splicing in the complexity of regulation of gene expression and protein expression, it is no longer sufficient to report up- or down-expression of genes and proteins without detecting and quantitating the splice isoforms. A variety of computational methods can be applied to deduce the functional significance of the sequence differences and folding differences in the protein splice isoforms, as we have demonstrated in this analysis of inflammatory breast cancer cell lines with contrasting pathways activated.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is supported by National Institutes of Health [1R21NS082212-01] and NIH [University of Michigan O'Brien Kidney Translational Core Center].

References

1. Zhang J, Manley JL. Misregulation of pre-mRNA alternative splicing in cancer. *Cancer Discov.* 2013; 3:1228–37. [PubMed: 24145039]
2. Paik YK, Omenn GS, Thongboonkerd V, Marko-Varga G, Hancock WS. Genome-wide proteomics, Chromosome-centric Human Proteome Project (C-HPP), part II. *J Proteome Res.* 2014; 13:1–4. [PubMed: 24328071]

3. Lane L, Bairoch A, Beavis RC, Deutsch EW, Gaudet P, Lundberg E, Omenn GS. Metrics for the Human Proteome Project 2013-2014 and strategies for finding missing proteins. *J Proteome Res.* 2014; 13:15–20. [PubMed: 24364385]
4. Omenn GS, Menon R, Zhang Y. Innovations in proteomic profiling of cancers: Alternative splice variants as a new class of cancer biomarker candidates and bridging of proteomics with structural biology. *J Proteomics.* 2013; 90:28–37. [PubMed: 23603631]
5. Omenn GS. Plasma proteomics, the Human Proteome Project, and cancer-associated alternative splice variant proteins. *Biochim Biophys Acta.* 2013; 1016:1016/j.bbapap.2013.10.016
6. Legrain P, Aebersold R, Archakov A, Bairoch A, Bala K, Beretta L, et al. The human proteome project: current state and future direction. *Mol Cell Proteomics.* 2011; 10:M111 009993. [PubMed: 21742803]
7. Paik YK, Jeong SK, Omenn GS, Uhlen M, Hanash S, Cho SY, et al. The Chromosome-centric Human Proteome Project for cataloging proteins encoded in the genome. *Nat Biotechnol.* 2012; 30:221–3. [PubMed: 22398612]
8. Marko-Varga G, Omenn GS, Paik YK, Hancock WS. A first step toward completion of a genome-wide characterization of the human proteome. *J Proteome Res.* 2013; 12:1–5. [PubMed: 23256439]
9. Liu S, Im H, Bairoch A, Cristofanilli M, Chen R, Deutsch EW, et al. A chromosome-centric human proteome project (C-HPP) to characterize the sets of proteins encoded in chromosome 17. *J Proteome Res.* 2013; 12:45–57. [PubMed: 23259914]
10. Zhang EY, Cristofanilli M, Robertson F, Reuben JM, Mu Z, Beavis RC, et al. Genome wide proteomics of ERBB2 and EGFR and other oncogenic pathways in inflammatory breast cancer. *J Proteome Res.* 2013; 12:2805–17. [PubMed: 23647160]
11. Menon R, Im H, Zhang EY, Wu SL, Chen R, Snyder M, Hancock WI, Omenn GS. Distinct splice variants and pathway enrichment in the cell-line models of aggressive human breast cancer subtypes. *J Proteome Res.* 2014; 13:212–27. [PubMed: 24111759]
12. Menon R, Omenn GS. Proteomic characterization of novel alternative splice variant proteins in human epidermal growth factor receptor 2/neu-induced breast cancers. *Cancer Res.* 2010; 70:3440–9. [PubMed: 20388783]
13. Dressman MA, Baras A, Malinowski R, Alvis LB, Kwon I, Walz TM, et al. Gene expression profiling detects gene amplification and differentiates tumor types in breast cancer. *Cancer Res.* 2003; 63:2194–9. [PubMed: 12727839]
14. Whiteaker JR, Zhang H, Zhao L, Wang P, Kelly-Spratt KS, Ivey RG, et al. Integrated pipeline for mass spectrometry-based discovery and confirmation of biomarkers demonstrated in a mouse model of breast cancer. *J Proteome Res.* 2007; 6:3962–75. [PubMed: 17711321]
15. Menon R, Zhang Q, Zhang Y, Fermin D, Bardeesy N, DePinho RA, et al. Identification of novel alternative splice isoforms of circulating proteins in a mouse model of human pancreatic cancer. *Cancer Res.* 2009; 69:300–9. [PubMed: 19118015]
16. Katz E, Dubois-Marshall S, Sims AH, Faratian D, Li J, Smith ES, et al. A gene on the HER2 amplicon, C35, is an oncogene in breast cancer whose actions are prevented by inhibition of Syk. *Br J Cancer.* 2010; 103:401–10. [PubMed: 20628393]
17. Grunewald TG, Kammerer U, Kapp M, Eck M, Dietl J, Butt E, et al. Nuclear localization and cytosolic overexpression of LASP-1 correlates with tumor size and nodal-positivity of human breast carcinoma. *BMC Cancer.* 2007; 7:198. [PubMed: 17956604]
18. Fagerberg L, Oksvold P, Skogs M, Algenas C, Lundberg E, Ponten F, et al. Contribution of antibody-based protein profiling to the Human Chromosome-centric Proteome Project (C-HPP). *J Proteome Res.* 2013; 12:2439–48. [PubMed: 23276153]
19. Cox J, Mann M. 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics.* 2012; 13(16):S12. [PubMed: 23176165]
20. Holbro T, Beerli RR, Maurer F, Koziczak M, Barbas CF 3rd, Hynes NE. The ErbB2/ErbB3 heterodimer functions as an oncogenic unit: ErbB2 requires ErbB3 to drive breast tumor cell proliferation. *Proc Natl Acad Sci U S A.* 2003; 100:8933–8. [PubMed: 12853564]

21. Menon R, Roy A, Mukherjee S, Belkin S, Zhang Y, Omenn GS. Functional implications of structural predictions for alternative splice proteins expressed in Her2/neu-induced breast cancers. *J Proteome Res.* 2011; 10:5503–11. [PubMed: 22003824]
22. Jin Q, Yuan LX, Boulbes D, Baek JM, Wang YN, Gomez-Cabello D, Hawke DH, Yeung SC, Lee MH, Hortobagyi GN, Hung MC, Esteva FJ. Fatty acid synthase phosphorylation: a novel therapeutic target in HER2-overexpressing breast cancer cells. *Breast Cancer Res.* 2010; 12:1–18.
23. Kumar-Sinha C, Ignatoski KW, Lippman ME, Ethier SP, Chinnaiyan AM. Transcriptome analysis of HER2 reveals a molecular connection to fatty acid synthesis. *Cancer Res.* 2003; 63:132–9. [PubMed: 12517789]
24. Houchens NW, Merajver SD. Molecular determinants of the inflammatory breast cancer phenotype. *Oncology.* 2008; 22:1556–61. discussion 61, 65–8, 76. [PubMed: 19198217]
25. Wu M, Wu ZF, Rosenthal DT, Rhee EM, Merajver SD. Characterization of the roles of RHOC and RHOA GTPases in invasion, motility, and matrix adhesion in inflammatory and aggressive breast cancers. *Cancer.* 2010; 116:2768–82. [PubMed: 20503409]
26. Eksi R, Li HD, Menon R, Wen Y, Omenn GS, Kretzler M, et al. Systematically differentiating functions for alternatively spliced isoforms through integrating RNA-seq data. *PLoS Comput Biol.* 2013; 9:e1003314. [PubMed: 24244129]
27. Oberwinkler J, Lis A, Giehl KM, Flockerzi V, Philipp SE. Alternative splicing switches the divalent cation selectivity of TRPM3 channels. *J Biol Chem.* 2005; 280:22540–8. [PubMed: 15824111]
28. Revil T, Toutant J, Shkreta L, Garneau D, Cloutier P, Chabot B. Protein kinase C-dependent control of Bcl-x alternative splicing. *Mol Cell Biol.* 2007; 27:8431–41. [PubMed: 17923691]
29. Vegrn F, Boidot R, Oudin C, Riedinger JM, Bonnetain F, Lizard-Nacol S. Overexpression of caspase-3s splice variant in locally advanced breast carcinoma is associated with poor response to neoadjuvant chemotherapy. *Clin Cancer Res.* 2006; 12:5794–800. [PubMed: 17020986]
30. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet.* 2008; 40:1413–5. [PubMed: 18978789]
31. Chung SS, Giehl N, Wu Y, Vadgama JV. STAT3 activation in HER2-overexpressing breast cancer promotes epithelial-mesenchymal transition and cancer stem cell traits. *Int J Oncol.* 2014; 44:403–11. [PubMed: 24297508]
32. Baron AT, Cora EM, Lafky JM, Boardman CH, Buenafe MC, Rademaker A, et al. Soluble epidermal growth factor receptor (sEGFR/sErbB1) as a potential risk, screening, and diagnostic serum biomarker of epithelial ovarian cancer. *Cancer Epidemiol Biomarkers Prev.* 2003; 12:103–13. [PubMed: 12582019]
33. Wilken JA, Baron AT, Maihle NJ. The epidermal growth factor receptor conundrum. *Cancer.* 2010; 117:2358–60. [PubMed: 24048781]
34. Center for Cancer Research. [Updated December 18, 2013. Accessed March 06, 2014] Our Science - Ried Website. <http://ccr.cancer.gov/staff/staff.asp?profileid=5680>
35. Misteli T. Beyond the sequence: cellular organization of genome function. *Cell.* 2007; 128:787–800. [PubMed: 17320514]
36. Roukos V, Voss TC, Schmidt CK, Lee S, Wangsa D, Misteli T. Spatial dynamics of chromosome translocations in living cells. *Science.* 2013; 341:660–4. [PubMed: 23929981]
37. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature.* 2011; 478:64–9. [PubMed: 21909114]
38. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, et al. The evolutionary landscape of alternative splicing in vertebrate species. *Science.* 2012; 338:1587–93. [PubMed: 23258890]
39. Merkin J, Russell C, Chen P, Burge CB. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science.* 2012; 338:1593–9. [PubMed: 23258891]
40. Kaida D, Motoyoshi H, Tashiro E, Nojima T, Hagiwara M, Ishigami K, et al. Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat Chem Biol.* 2007; 3:576–83. [PubMed: 17643111]

41. Papasaikas P, Valcarcel J. Evolution. Splicing in 4D. *Science*. 2012; 338:1547–8. [PubMed: 23258879]

Highlights

- Alternative splicing is a fascinating evolutionary development of multi-exon genes.
- ERBB2+ and EGFR+ breast cancer cell lines have very different splicing and pathways.
- We relate splicing patterns to interaction networks and signaling pathways.
- We review our methods for predicting functional variants among isoforms.
- Understanding upstream ligands and downstream pathways may guide therapy.

Figure 1.

The bar chart shows Reads Per Kilobase per Million [RPKM] for splice variant transcripts of ERBB2 (Figure 1a) and for CDK12, FBXL20, and GRB7 (Figure 1b) which are part of the ERBB2 amplicon. In Figure 1a, SUM190 has remarkably high expression of variant ERBB2_1225 (ENSP00000385185), with much lower, but detectable reads (RPKM 1-2) for this variant detected in four of the other five cell lines. Figure 1b shows highly differential expression of splice variants of each of three genes (CDK12(ENSP00000398880, 1490 aa), FBXL20 (ENSP00000264658, 436 aa), and GRB7(ENSP00000310771, 532 aa) that are part of the ERBB2 amplicon at 17q12. (Reprinted with permission from Liu, S., et al. (2013). "A chromosome-centric human proteome project (CHPP) to characterize the sets of proteins encoded in chromosome 17." *J Proteome Res* 12: 45-57. Copyright 2013 American Chemical Society.)

Figure 2.

Heat map showing the total distinct RNA Seq reads for the transcripts of genes from 17q12 region identified in SKBR3, SUM190 and SUM149. (Reprinted with permission from Menon et al. (2013). "Distinct splice variants and pathway enrichment in the cell-line models of aggressive human breast cancer subtypes." *J Proteome Res* 13: 212-27. Copyright 2013 American Chemical Society.)

Figure 3.

A composite of SUM149 (A) and SUM190 (B) transcriptomic, proteomic, and interaction data for significant oncogenes observed in SUM149 and SUM190. The following notations are used. Line length: Interaction score (shorter line, stronger interaction with ERBB2). Circle size: RPKM value (largest: RPKM > 15, medium: RPKM between 3 and 15, small: RPKM between 1 and 3, spot: RPKM < 1). Black circle: if observed in proteomic experiments. Percentage: percentage of proteins identified in SUM149 or 190 with specific oncogene interactions as listed by STRING or I2D in Genecards.org. (Reprinted with permission from EY Zhang et al. (2013). "Genome wide proteomics of ERBB2 and EGFR and other oncogenic pathways in inflammatory breast cancer." *J Proteome Res* 12: 2805-17. Copyright 2013 American Chemical Society.)

Figure 4.

The relative expression levels based on total distinct reads identified for the transcripts. Seven splice variants of UBC were identified. One variant each was unique to SKBR3 and to SUM149 and two were unique to SUM190. The canonical variant UBC_685 was expressed in all three cancer cell lines (Reprinted with permission from Menon et al. (2013). “Distinct splice variants and pathway enrichment in the cell-line models of aggressive human breast cancer subtypes.” *J Proteome Res* 13: 212-27. Copyright 2013 American Chemical Society.)

Figure 5.

The relative expression levels based on total distinct reads identified for the transcripts. Five splice variants of STAT3 were identified. Two variants were unique to SKBR3. (Reprinted with permission from Menon et al. (2013). "Distinct splice variants and pathway enrichment in the cell-line models of aggressive human breast cancer subtypes." *J Proteome Res* 13:212-27. Copyright 2013 American Chemical Society.)

Table 1
Summary of the Features of Breast Cancer Cell Lines SKBR3, SUM190, SUM 149

Feature	SKBR3	SUM190	SUM149
Origin	Established line	Primary tumor line	Primary line
Histology	Adenocarcinoma ER-/PR-, metastatic	Inflammatory BC ER-/Pr-	Inflammatory BC ER-/Pr-
ERBB2 (HER2/neu)	300	400	14
ERBB1 (EGFR)	1.4	n.d.	60