**MODULE 1 – Basics of breast cancer**

**Know someone that has or had breast cancer?**

***Most likely you do!*** **In 2024, there were** [**300,000 diagnosed**](https://www.breastcancer.org/facts-statistics) **cases of breast cancer.** As you may know, cancer is the outcome of uncontrolled cell division. Cancerous cells have lost the ability to control the pathway leading to cell division and the compromised cell’s division continues unabated to potentially form a tumor and possibly migrate beyond the original site.  The genes that regulate these processes of cell division which have lost function are known as tumor suppressor genes, and the ones that may be increased in function are known as [oncogenes](https://www.genome.gov/genetics-glossary/Oncogene#:~:text=An%20oncogene%20is%20a%20mutated,in%20regulating%20normal%20cell%20division.).

A graph of pink bars

Description automatically generated

**Figure 1.** Breast cancer incidence stratified by race and ethnicity. Taken from <https://www.komen.org/breast-cancer/facts-statistics/breast-cancer-statistics/>

***What causes breast tissue to become cancerous?***The major risk factor for breast cancer is age. Most breast cancers are diagnosed in women over 50. The accumulation of DNA damage with age and lifestyle may result in errors in the cellular division genes discussed above. The BRCA-1 and BRCA-2 genes that are famously associated with with breast cancer actually govern cellular DNA repair mechanisms, so the if the ability to repair DNA damage is compromised due to mutation in BRCA-1 gene, DNA damage within the cell division genes may accumulate to the point of change in function of the genes and therefore uncontrolled cell division or cancer.

To diagnose a patient with breast cancer, there is a tremendous amount of testing to give the physician (and the patient!) a clear, in-depth picture of all the information about their patient’s cancer. Patients endure numerous tests- PET scans, ultrasounds, MRIs, bloodwork, and others to give the physician the clearest understanding of the physical traits of the cancer cell and therefore how to attack the cancer.  You might have heard of “**triple negative”** or **“triple positive”** subtypes of breast cancer.  These cancer subtypes correspond to several genes that may or may not be expressed in the patient's tumor.  The terms triple negative and triple positive specifically identify estrogen receptor (ER), progesterone receptor (PR), and the Her2 activity.  The presence or absence of these factors represent a pivotal point in the diagnosis of breast cancer.  The profile of these factors generally determine treatment protocols.

**Review the following links to better understand clinical cancer subtypes:**

**Triple negative-** [**Penn Medicine**](https://www.pennmedicine.org/cancer/types-of-cancer/breast-cancer/types-of-breast-cancer/triplenegative-breast-cancer#:~:text=If%20the%20cancer%20cells%20test,as%20triple%20negative%20breast%20cancer.)

**Triple positive-** [**NIH National Cancer Institute**](https://www.cancer.gov/publications/dictionaries/cancer-terms/def/triple-positive-breast-cancer)

* **Based on your review of the links presented above, what is the recommended treatment protocol for each type of breast cancer?**

**MODULE 2 – The story of Anna and Betty**

It has been a busy day for Dr. Ferreiro, a breast cancer specialist at NIH Cancer Hospital. Her first patient for the day was Anna. Anna is a 50-year-old Caucasian American (CA) woman with hypertension which is managed by medication. She has no family history of breast cancer. She does not smoke, leads a sedentary lifestyle due to her desk job and uses alcohol occasionally. Her symptoms included a persistent breast lump with occasional pain but no significant weight loss. Dr. Ferreiro ordered an initial biopsy to assess the tumor progression. The result came back as moderate grade TNBC, Stage III with no lymph involvement.

* **How is the stage and grade of a tumor determined?**
* **What are the key characteristics of Triple Negative Breast Cancer (TNBC), and how do these characteristics impact treatment options?**

The next patient in Dr. Ferreiro’s office was Betty, a 48-year-old African American (AA) woman with no significant comorbidities, who reports an otherwise healthy lifestyle. She reports that her mother passed away from breast cancer 12 years ago. She is a non-smoker, exercises regularly, and moderately uses alcohol. She complains of fatigue, localized pain in the breast, and mild weight loss.

The result diagnosed Betty with high grade triple negative breast cancer (TNBC), Stage III, 6 cm with lymph node involvement

* **How do the tumor grade and stage of Betty’s and Anna’s cancers compare, and what implications might these differences have for their treatment plans?**
* **Considering Betty's lymph node involvement and Anna's lack of lymph node involvement, how might these factors influence their treatment strategies?**

The oncologist decided to start a treatment based on cisplatin, a chemotherapy drug that damages the DNA of cancer cells, leading to cell death.

**MODULE 3 – Follow up appointments**

After six months of treatment, both women return for a follow-up. Unfortunately, they continue to exhibit symptoms. Betty’s case is worsened compared to Anna, but neither of the cases shows any significant improvement, indicating that cisplatin alone is not effectively managing their disease. Given the persistence of their tumor progression, the oncologist decides to explore alternative therapies. After careful discussion with Dr. Tripathi, a molecular oncology colleague, Dr. Ferreiro decides to order additional testing of the biopsy to check for androgen receptor (AR) immunopositivity on the tumor samples. The decision is rooted in emerging evidence that AR signaling may contribute to breast cancer progression, especially in cases where standard chemotherapy, like cisplatin, fails.

The AR staining (Figure 2)reveals that both Anna’s and Betty’s tumors are AR-positive, meaning that they express androgen receptors. With this new information, the oncologist decides to prescribe enzalutamide, an AR antagonist, to both Anna and Betty.

A close-up of a microscope slide

Description automatically generated

**Figure 2.** Immunohistochemical Staining for androgen receptor (AR) and haematoxylin-eosin and saffron (HES) in two cases. A-B: Case number 4046 (Betty). A, AR-positive nuclear staining in ≥ 10%, HES. B, Control AR-negative. C-D: Case number 4067 (Anna). C, AR-positive nuclear staining in ≥ 10%, HES. D, Control AR-negative.

**·       What is the role of androgen receptors (AR) in breast cancer progression, and why might AR expression be relevant in the context of TNBC that is not responding to standard chemotherapy?**

**·       How does enzalutamide, an AR antagonist, work to inhibit cancer cell growth, and why might it be a suitable treatment option for patients with AR-positive TNBC?**

**·       What are the mechanisms by which AR signaling can contribute to breast cancer progression, particularly in cases where traditional chemotherapy has been ineffective?**

**INSTRUCTOR NOTE:**

**Mechanism of Action:** Enzalutamide works by binding to the androgen receptor, preventing androgens (male hormones like testosterone) from binding and activating the receptor. Normally, when androgens bind to AR, the receptor translocates to the nucleus and promotes the expression of genes that drive cell proliferation and survival. By blockifeng this pathway, enzalutamide inhibits the growth of AR-positive cancer cells.

With renewed hope, both Betty and Anna begin their new treatment. They return for their next follow-up after another 6 months. This time, the outcomes diverge. Anna shows significant improvement; her cancer appears to be in remission, with no signs of recurrence. However, Betty’s cancer seems to have worsened, leading to a poor prognosis.

As an oncologist, Dr. Ferreiro is now faced with a critical question: Why did the AR-targeted therapy work for one patient but not the other? Given this unexpected result, she considers that there may be underlying biological differences in how each patient’s cancer expresses or responds to androgen receptor signaling.

* **What factors could explain the significant difference in response to enzalutamide between Anna and Betty? (Consider differences in tumor biology, AR expression levels, previous treatment responses, and individual patient factors).**
* **What are the possible reasons for the recurrence and worsening of Betty’s cancer despite AR-targeted therapy? (Consider potential mechanisms of resistance to enzalutamide, changes in tumor biology, or secondary mutations).**

Betty’s case has Dr. Ferreiro stumped. She needs to investigate it further. Following Dr. Tripathi’s advice she orders genetic sequencing of the AR gene.

* **Given that both patients are AR-positive, but only one responded well to enzalutamide, what possible genetic or molecular mechanisms could explain the difference in treatment outcomes? Consider factors like mutations in the AR gene, alternative splicing, or differences in downstream signaling pathways.**

To avoid similar problems in the future, the oncologist decides to use more advanced diagnostic techniques, such as Immunohistochemistry (IHC) and Western blotting, to analyze protein expression and function in the tumor tissues. These methods allow for a more detailed assessment of the AR pathway’s reliability, helping to tailor treatments more effectively from the outset.

**Summarize the data from both patients by filling in the following table:**

Table 1:

The results from the sequencing can be found in ENTER FILE LINK HERE

**MODULE 4 – Molecular analysis of patient gene**

To analyze Betty’s cDNA sequence, follow the steps below:

1. Go to [BLAST](https://blast.ncbi.nlm.nih.gov/Blast.cgi) and click Nucleotide BLAST

A screenshot of a computer

Description automatically generated

1. Copy the FASTA sequence from the end of this document and paste the sequence (along with the header) into the blastn Query Sequence box. Using all of the default search parameters, click BLAST to begin the search.

A screenshot of a search box

Description automatically generated

1. The default search parameters will produce 100 search results that are very similar. *Which results would be considered the “best hit”? How do you determine which results would be considered the best?*A screenshot of a computer

   Description automatically generated
2. To narrow down the search results to give you the best results, rerun your search using the RefSeq Select database and specifying the organism. Leave the remaining search parameters as default.

A screenshot of a computer

Description automatically generated

1. Using the results from your more advanced search, click the alignment tab and set the alignment view as “Pairwise with dots for identities”.

A screenshot of a computer

Description automatically generated

1. Are there differences between your query (the sequence you input) and the subject (the sequence that your input is aligned to)? List any differences you find in Table 2 below.

**Table 2:**

|  |  |  |  |
| --- | --- | --- | --- |
| Nucleotide position of difference for query | Query sequence nucleotide | Nucleotide position of difference for subject | Subject sequence nucleotide |
|  |  |  |  |
|  |  |  |  |

1. From this page, you can determine the conserved domains (regions of the protein that carry out a specific function) of the androgen receptor. To find these domains, click on the GenBank link.
2. Scroll down on the GenBank page and click on Protein until the Related information on the right-hand side of the page.A screenshot of a computer

   Description automatically generatedA screenshot of a computer

   Description automatically generated
3. On the right side of the protein page, click Identify Conserved Domains.A screenshot of a computer

   Description automatically generated

A screenshot of a computer

Description automatically generated

1. What conserved domains are found in the androgen receptor? What is the range of amino acids for each domain?

Return to your BLAST results and look at the sequence alignment tab. Are there differences between your query (the sequence you input) and the subject (the sequence that your input is aligned to)? List any differences you find in Table 1 below.

**Table 1:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nucleotide position of difference for query | Query sequence nucleotide position | Nucleotide position of difference for subject | Subject sequence nucleotide position | In which protein domain is the mutation found? |
|  |  |  |  |  |
|  |  |  |  |  |

To explore what the effects of these mutations are, you will use a tool to translate the patient nucleotide sequence to her protein sequence.

1. Visit [ORFfinder](https://www.ncbi.nlm.nih.gov/orffinder/), a web-based tool that detects all possible open reading frames (ORFs) in a given nucleotide sequence.
2. Paste the patient’s nucleotide FASTA sequence into the entry box. Scroll down and check the box next to the “Ignore nested ORFs” option.

INSTRUCTOR NOTE:

The description line of the FASTA file can be included or omitted in the search box of the ORFfinder tool without affecting the performance of the tool.

1. Take a close look at the optional search parameters available in the tool. You can specify the minimal length of the ORF, the type of start codon to use, and whether to ignore nested ORFs.

**What property of the genetic code comes to mind when you see the options in the dropdown menu for the genetic code to be used?**

INSTRUCTOR NOTE: This last question is meant to remind the students that the genetic code is *nearly* universal, but the translation of open reading frames in certain organisms or organelles (like mitochondria or some protists) may result in different gene products. If you opt to omit ignoring the nested ORF, there are 20 ORFs in the results; the appropriate result to continue will be ORF1. This will slightly change the workflow for the next few steps.

1. Submit the job with all other parameters left in the default options.
2. In the results page, a genome viewer indicates the position of the identified ORFs.

A screenshot of a computer

Description automatically generated

Since you chose to ignore all smaller (nested) ORFs, there is only one ORF (termed ORF1) shown within the viewer window. Its sequence is shown in a separate window to the lower left part of the page

1. Click on the ‘Mark’  button. This allows you to download the protein sequence as a FASTA file. To do that, click on the blue ‘Download Marked Set’ button to the right.
2. Save the protein FASTA file with an appropriate name, e.g. *AR\_patient.fasta*.
3. The next step is to see how the patient-obtained patient sequence is different from the wild-type protein. ORFfinder gives you the opportunity to perform a BLAST search directly from the tool.
4. Change the database to the bottom left of the screen to ‘Reference Proteins (refseq\_protein)’ and click on the BLAST button.

A close-up of a yellow and black sign

Description automatically generated

1. Click on the blue BLAST button.

A blue rectangular buttons with white text

Description automatically generated

INSTRUCTOR NOTE: The SmartBLAST option will search the non-redundant (nr) protein database with an optimized version of BLAST targeted to closely related sequences. This is not the desired method in this case, as the focus is to align the sequence obtained from our patient to the reference human protein sequence so that we can make a comparison.

1. Look over the BLAST results page.

**Do your results reflect what you expected from the search?**

1. Click on the ‘Alignments’ tab of the results page.
2. In the ‘Alignment View’ banner, select ‘Pairwise with dots for identities’ from the dropdown menu, so that the differences between the query (the patient-derived sequence) and the subject (the database-deposited sequence) are most obvious.

**What effect do the mutation(s) observed from this patient has/have on the gene product?**

INSTRUCTOR NOTE: This question can be as broad or focused as needed, and it can be modified to be used as a multiple-choice question, fill-in-the-blank, and other types. The answer here is T575A and T878A.

**MODULE 5 – Exploration of the mutant protein structure**

Now that Dr. Ferreiro knows what the changes are in Berry’s AR protein, she wants you to explore the structure.

1. Go to [iCn3D](https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html), a web-based structure viewer, type in AR’s PDB ID “2AM9”, and hit enter, or click “load biological unit”. This structure is a crystal structure of human androgen receptor ligand binding domain in complex with testosterone, published in the PDB [here](https://www.rcsb.org/structure/2am9).

*Alternatively, click the link from Blastp from the previous step to get to iCn3D structure.*

Get familiar with iCn3D basic functions such as define sets, changing color and style, select, etc., with this [iCn3D Fundamentals](https://www.nlm.nih.gov/ncbi/workshops/2024-07_ISMBHackathon-3d-molecular-structures/icn3d-fundamentals.html) and [Help Docs](https://www.ncbi.nlm.nih.gov/Structure/icn3d/docs/icn3d_help.html).

[Follow](https://structure.ncbi.nlm.nih.gov/icn3d/share.html?AHyFbB74Tb8WYrAn6) the instructions in the link (color to rainbow, side chain to line, opened defined sets, and seq and annotations)

1. Find and highlight the T877 in Sequence and Annotations.

INSTRUCTOR NOTE: depending on the source of information, it might be T878. Emphasize to students that do not depend solely on numbering, but they need to check the nearby sequences as well.

[Follow](https://structure.ncbi.nlm.nih.gov/icn3d/share.html?RKy1jgVd1wgWdCih7) the instructions in the link (whole structure color by atom, and TES to magenta to see mutations easier. T877 color to green, side chain to stick)

1. Check what T877 is interacting with, by Select by Distance

A screenshot of a computer

Description automatically generated

After clicking Display, the testosterone molecule is clearly highlighted, showing interaction between T877 and TES.

Now you can mutate T877 to alanine and see how the mutation potentially affects the interaction with testosterone.

INSTRUCTOR NOTE: this function only changes the side chain of the particular amino acid in the structure, but does not explain how this would affect protein folding *in cellulo*.

A screenshot of a computer

Description automatically generated

It would pop out a [new page](https://structure.ncbi.nlm.nih.gov/icn3d/share.html?216Ggv6w6SHXEZas6), and pressing the letter ‘a’ would alternate between WT and mutant. Note: you may need to change the style for better visualization.

A screenshot of a computer game

Description automatically generated

A computer screen shot of a molecule

Description automatically generated

1. Now back to the 2AM9 structure. In the Sequence and Annotations window, turn on ClinVar, and check out the different clinically important SNPs. Pick one ClinVar entry of interest (except T877A), explore its location on the structure, compare the difference between WT and mutant, check the ClinVar links for further information, and make hypothesis: what is the impact of the mutation and why it potentially leads to disease. A screenshot of a computer

   Description automatically generated

**What is the accession ID for the T877A mutation in ClinVar? What is the accession ID in the dbSNP database?**

**MODULE 6 – How can basic science inform the patient outcome?**

1. In the previous module, you worked through the structural analysis of the T877A mutation in Betty’s AR gene using the human ligand binding domain structure (PDB ID 2AM9). From your BLAST analysis, you found a second mutation. What additional mutation did you find? In which protein domain is this other mutation found?
2. Unlike the ligand binding domain, there is no solved crystal structure of the DNA binding domain for AR. However, there is a structure available for a homolog of AR from the Norwegian Rat bound to DNA (PDB ID 1R4I). Using the FASTA protein sequence output from ORF finder in Module 4, use blastp to align the sequence to the rat version of the protein to Betty’s AR sequence. Do these sequences align well? How can you tell? Does the T575 from the human version of AR align with the rat version?
3. Analyze the 1R4I structure in iCn3D. Where is the mutation in the DNA binding domain found within the rat structure? What does it interact with? What would you predict would happen if you mutated this threonine to an alanine?
4. In 2006, Monge et al. researched the functional significance of the T575A and T877A in the human AR protein and the relation of these mutations to prostate cancer. Using the data presented in the paper (point them to specific data?) and your iCn3D analysis of the rat version of the DNA binding domain, predicted significance of T575A and how does this analysis connect to Betty’s breast cancer prognosis/treatment?