1. **Using data from:**

* Molecular Linkage file (main file)
* Diagnosis file (for primary site and stage info)
* Metastatic Disease (to help assign stage)

1. **Definitions**

* **Melanoma diagnosis** (Diagnosis file): HistologyCode = {list of melanoma codes previously sent}
* **Tumor sequenced** (Molecular Linkage file): Tumor/Germline variable = Tumor

**New variables to create**

* **3. MelanomaDiagnosisCount:** count for number of melanoma clinical diagnoses for a patient
  + Create using the number of unique [AgeAtDiagnosis and PrimaryDiagnosisSite combinations] for each patient
  + This approach may work best since a few patients have multiple melanomas diagnosed at same age, so it cannot be used on its own.
    - *Example with multiple diagnoses at same age with same stage: ILE2DL0KMW*
* **4. SequencedTumorCount**: count for number of sequenced tumor samples for a patient
  + Create using the number of unique [DeidSpecimenID and AvatarKey combinations] for each patient
  + This approach may work best since a few patients have multiple sequenced tumors at the same age (or stage, etc), so these variables cannot be used on their own.
    - *Example with multiple tumors sequenced at same age/stage: 59OP5X1AZL*
* **5. AssignedPrimarySite**: {cutaneous, ocular, mucosal, unknown}
  + Based on the parameters outlined below; this will be the primary site variable used for the analysis
* **6. AssignedStage**: {I, II, III, IV}
  + Based on the parameters outlined below; this will be the stage variable used for the analysis

1. **7. Using the definitions above, identify the following patient groups:** 
   * **Group A**: **MelanomaDiagnosisCount** = 1 AND **SequencedTumorCount =** 1
     + (n=327)
   * **Group B**: **MelanomaDiagnosisCount** = 1 AND **SequencedTumorCount >** 1
     + (n=19)
   * **Group C**: **MelanomaDiagnosisCount** > 1 AND **SequencedTumorCount =** 1
     + (n=30)
   * **Group D**: **MelanomaDiagnosisCount** > 1 AND **SequencedTumorCount >** 1
     + (n=3)
2. **Start with Group A** (1 melanoma diagnosis and 1 tumor sequenced)
   * **First, identify the patients with primary cutaneous melanoma** 
     + IF PrimaryDiagnosisSite contains “skin” OR “ear” OR “eyelid” OR “vulva”, THEN **AssignedPrimarySite** = cutaneous
     + *Vulvar melanoma is included here given that all appear to have been staged as cutaneous (not mucosal) melanoma.*
   * **Next, assign the stage for these patients using the following rules. All of these rules assume AssignedPrimarySite = cutaneous.**

* + - IF PathGroupStage contains “IV”, then **AssignedStage** = “IV”
    - IF ClinGroupStage contains “IV” AND PathGroupStage is [“Unknown/Not Reported” OR “No TNM applicable for this site/histology combination” OR “Unknown/Not Applicable”], THEN **AssignedStage** = “IV”

*NOTES: The above two rules define the patients with stage IV disease at initial diagnosis, so the specimen is stage IV regardless of the time interval between diagnosis and specimen collection.*

* + - * *This assumes all specimens are obtained after initial diagnosis. This is true for all patients except one patient in Group B (so will be dealt with by other rules in that section).*
      * *There is a small chance that a few specimens obtained after initial diagnosis of stage IV disease may instead represent new primary melanomas, but this is not able to be confirmed with the information in our dataset and these rules align with the most common, expected scenario.*
      * *For stage IV, “contains” is used instead of equals since a few have stage IVA, IVB, or IVC. These substages are used for mucosal melanoma of head and neck. The ones captured by this rule appear to be cutaneous (skin of scalp and neck; skin of trunk), so will be treated as stage IV cutaneous.*
    - IF SpecimenSiteofCollection *does not contain* [“skin” OR “lymph node” OR “soft tissues” OR “muscle” OR “parotid” OR “chest wall” OR “head”] AND Primary/Met = “Metastatic”, THEN **AssignedStage** = “IV”

*NOTES: This rule defines patients that have a tumor specimen from a metastatic site that is not a cutaneous or nodal site. These represent current stage IV disease at age (time) of specimen collection, regardless of the associated staging information at the age (time) of initial diagnosis reported in the Diagnosis file.*

* + - IF SpecimenSiteofCollection contains “lymph node” AND [MetsDzPrimaryDiagnosisSite = “Unknown/Not Applicable” OR {combined entry on Metastatic file with: [MetsDzPrimaryDiagnosisSite contains [“skin” OR “ear” OR “eyelid” OR “vulva”] AND MetastaticDiseaseInd = [“Yes – Regional” OR “Yes – NOS”] AND MetastaticDiseaseSite contains “lymph node”]}] AND PathGroupStage contains “III”, THEN **AssignedStage** = “III”

*Examples: 2AP9EDU231, 300QLLH1JW*

*Examples to confirm with Slingluff: QLWU5QNQIB, DTUPUJ06B5*

*NOTES: This rule defines patients that had stage III (node positive) disease at diagnosis, and have a nodal tumor specimen most likely from either the regional nodal disease at initial diagnosis or recurrent regional nodal disease. Both scenarios result in stage III disease; however, we cannot assign a substage (III: A,B,C,D) without additional information or rules, but this should be fine for the purposes of our analysis.*

* + - IF SpecimenSiteofCollection contains “lymph node” AND [MetsDzPrimaryDiagnosisSite = “Unknown/Not Applicable” OR {combined entry on Metastatic file with: [MetsDzPrimaryDiagnosisSite contains [“skin” OR “ear” OR “eyelid” OR “vulva”] AND MetastaticDiseaseInd = [“Yes – Regional” OR “Yes – NOS”] AND MetastaticDiseaseSite contains “lymph node”]}] AND {PathGroupStage does *not* contain [“III” OR “IV”] OR ClinGroupStage does *not* contain “IV”}, THEN **AssignedStage** = “III”

*Example: 0VZ1LP53XJ*

*NOTES: This rule defines patients that had early stage (node negative) disease at diagnosis, and have a nodal tumor specimen most likely from recurrent disease in the regional nodal basin, resulting in stage III disease.* *We cannot assign a substage (III: A,B,C,D) without additional information or rules, but this should be fine for the purposes of our analysis*

* + - IF SpecimenSiteofCollection contains “lymph node” AND {combined entry on Metastatic file with: MetsDzPrimaryDiagnosisSite contains [“skin” OR “ear” OR “eyelid” OR “vulva”] AND MetastaticDiseaseInd = “Yes – Distant” AND MetastaticDiseaseSite contains “lymph node”} AND {PathGroupStage does *not* contain “IV” OR ClinGroupStage does *not* contain “IV”}, THEN **AssignedStage** = “IV”

*Example: 9TUP5F9XV7*

*NOTES: This rule defines patients who did not have an initial diagnosis of stage IV that have a tumor specimen most likely from a distant nodal recurrence that would be now considered stage IV disease.*

* + - IF SpecimenSiteofCollection contains ‘parotid’ AND {combined entry on Metastatic file with: MetsDzPrimaryDiagnosisSite contains [“skin” OR “ear” OR “eyelid” OR “vulva”] AND MetastaticDiseaseInd = “Yes – Distant” AND MetastaticDiseaseSite contains “parotid”} AND {PathGroupStage does *not* contain “IV” OR ClinGroupStage does *not* contain “IV”}, THEN **AssignedStage** = “IV”

*Example: EQ8CK21XQT*

*Example to confirm with Slingluff: JEGI7KSPVC*

*NOTES: This rule defines patients who did not have an initial diagnosis of stage IV that have a parotid tumor specimen that is most likely a distant metastatic recurrence.*

* + - IF SpecimenSiteofCollection contains ‘parotid’ AND [MetsDzPrimaryDiagnosisSite = “Unknown/Not Applicable” OR {combined entry on Metastatic file with: [MetsDzPrimaryDiagnosisSite contains [“skin” OR “ear” OR “eyelid” OR “vulva”] AND MetastaticDiseaseInd = [“Yes – Regional” OR “Yes – NOS”] AND MetastaticDiseaseSite contains “parotid” OR “lymph node”]}], THEN **AssignedStage** = “III”

*Example: GONDO7FLPP*

*NOTES: This rule defines patients that have a parotid tumor specimen that most likely represents regional nodal disease (stage III) from initial diagnosis. The MetastaticDiseaseSite in this case may contain lymph node instead of parotid (which may have been resected at time of neck dissection for regional nodes).*

* + - IF SpecimenSiteofCollection contains [“skin” OR “ear” OR “eyelid” OR “head” OR “soft tissues” OR “muscle” OR “chest wall” OR “vulva”] AND Primary/Met = “Metastatic” AND {PathGroupStage does *not* contain “IV” OR ClinGroupStage does *not* contain “IV”} AND {combined entry on Metastatic file with: MetsDzPrimaryDiagnosisSite contains [“skin” OR “ear” OR “eyelid” OR “vulva”] AND MetastaticDiseaseInd = “Yes – Distant” AND MetastaticDiseaseSite contains [“skin” OR “ear” OR “eyelid” OR “head” OR “soft tissues” OR “muscle” OR “chest wall”]} OR {combined entry on Metastatic file with: MetsDzPrimaryDiagnosisSite contains [“skin” OR “ear” OR “eyelid” OR “vulva”] AND MetastaticDiseaseInd = [“Yes – Distant” OR “Yes – NOS”] AND MetastaticDiseaseSite *does not contain* [“skin” OR “ear” OR “eyelid” OR “head” OR “soft tissues” OR “muscle” OR “chest wall” OR “vulva”] AND [AgeAtMetastaticSite <= AgeAtSpecimenCollection OR AgeAtMetastaticSite = “Age Unknown/Not Recorded”]}], THEN **AssignedStage** = “IV”

*Examples: 7EE85P5EGN, 9CJCPWLERP, ABE47J13C5, CG6JRI0XIX*

*Will try to edit this rule into multiple simpler rules if able*

*NOTES: This rule defines patients who did not have an initial diagnosis of stage IV that have a tumor specimen most likely from a distant cutaneous recurrence that would now be considered stage IV disease OR any cutaneous recurrence (including local/regional) after interval development of stage IV disease (at any distant site).*

* + - IF SpecimenSiteofCollection contains [“skin” OR “ear” OR “eyelid” OR “head” OR “soft tissues” OR “muscle” OR “chest wall” OR “vulva”] AND Primary/Met = “Metastatic” AND {PathGroupStage does *not* contain “IV” OR ClinGroupStage does *not* contain “IV”}, THEN **AssignedStage** = “III”

*Examples: 0BUVFK9AQ1, 6DL054517A, A594OFU98I*

*NOTES: This rule defines patients who did not have an initial diagnosis of stage IV that have a tumor specimen most likely from a regional cutaneous recurrence or the primary cutaneous lesion with node positive disease at initial diagnosis.*

* + - IF SpecimenSiteofCollection contains [“skin” OR “ear” OR “eyelid” OR “head” OR “soft tissues” OR “muscle” OR “chest wall” OR “vulva”] AND Primary/Met = “Primary” AND {PathGroupStage does *not* contain “IV” OR ClinGroupStage does *not* contain “IV”} AND [{combined entry on Metastatic file with: MetsDzPrimaryDiagnosisSite contains [“skin” OR “ear” OR “eyelid” OR “vulva”] AND MetastaticDiseaseInd = “Yes – Distant” AND [AgeAtMetastaticSite <= AgeAtSpecimenCollection OR AgeAtMetastaticSite = “Age Unknown/Not Recorded”]}, THEN **AssignedStage** = “IV”

*Examples: 2T0D2KKDXB*

*Examples to confirm with Slingluff: SJUY8UK34Q, TIZXXXVCV9*

*NOTES: This rule defines patients who did not have an initial diagnosis of stage IV that have a cutaneous tumor specimen after interval development of stage IV disease (at any distant site).*

* *There is a small chance that a few specimens obtained after the interval development of stage IV disease may instead represent a new primary melanoma.* 
  + - IF SpecimenSiteofCollection contains [“skin” OR “ear” OR “eyelid” OR “head” OR “soft tissues” OR “muscle” OR “chest wall” OR “vulva”] AND Primary/Met = “Primary” AND {PathGroupStage does *not* contain “IV” OR ClinGroupStage does *not* contain “IV”} AND [MetsDzPrimaryDiagnosisSite = “Unknown/Not Applicable” OR {combined entry on Metastatic file with: MetsDzPrimaryDiagnosisSite contains [“skin” OR “ear” OR “eyelid” OR “vulva”] AND MetastaticDiseaseInd = [“Yes – Regional” OR Yes – NOS] AND [AgeAtMetastaticSite <= [(90/365.25) + AgeAtSpecimenCollection] OR AgeAtMetastaticSite = “Age Unknown/Not Recorded”]} OR {combined entry on Metastatic file with: MetsDzPrimaryDiagnosisSite contains [“skin” OR “ear” OR “eyelid” OR “vulva”] AND MetastaticDiseaseInd = “Yes – Distant” AND AgeAtMetastaticSite > AgeAtSpecimenCollection }] AND does NOT have any {combined entry on Metastatic file with: MetsDzPrimaryDiagnosisSite contains [“skin” OR “ear” OR “eyelid” OR “vulva”] AND MetastaticDiseaseInd = “Yes – Distant” AND AgeAtMetastaticSite <= AgeAtSpecimenCollection }], THEN **AssignedStage** = *numerical* value of PathGroupStage OR ClinGroupStage (if PathGroupStage is [“Unknown/Not Reported” OR “No TNM applicable for this site/histology combination” OR “Unknown/Not Applicable”])

*Examples: 9HA9MZCSU2, P1GNI4I652, TZFPKMZP2K, EKZGU61JTP, 7IHUS604WY*

*Will try to edit this rule into multiple simpler rules if able*

*NOTES: This rule defines patients without stage IV disease who have a cutaneous tumor specimen that most likely represents the cutaneous lesion from the initial diagnosis or local recurrence.*

* *A few specimens likely represent new primary melanomas.*
  + **Define the patients with primary ocular melanoma** 
    - If PrimaryDiagnosisSite contains “choroid” OR “ciliary body” OR “conjunctiva”, then **AssignedPrimarySite** = ocular
    - **Assign the stage for these patients using the following rules. All of these rules assume AssignedPrimarySite = ocular. There are no primary ocular specimens in this dataset, which is not unexpected since treatment of primary lesion is usually radiation.** 
      * IF PathGroupStage contains “IV”, then **AssignedStage** = “IV”
      * IF ClinGroupStage contains “IV” AND PathGroupStage is [“Unknown/Not Reported” OR “No TNM applicable for this site/histology combination” OR “Unknown/Not Applicable”], THEN **AssignedStage** = “IV”

*NOTES: The above two rules define the patients with stage IV disease at initial diagnosis, so the specimen is stage IV regardless of the time interval between diagnosis and specimen collection.*

* + - IF SpecimenSiteOfOrigin contains [“eye” OR “choroid” OR “ciliary body” OR “conjunctiva”] AND {PathGroupStage does *not* contain “IV” OR ClinGroupStage does *not* contain “IV”} AND [{combined entry on Metastatic file with: MetsDzPrimaryDiagnosisSite contains “eye” OR “choroid” OR “ciliary body” OR “conjunctiva”] AND MetastaticDiseaseInd = “Yes – Distant” AND [AgeAtMetastaticSite <= AgeAtSpecimenCollection OR AgeAtMetastaticSite = “Age Unknown/Not Recorded”]}, THEN **AssignedStage** = “IV”

*Example: 2DC14Y2AQS*

*NOTES: This rule defines patients who did not have stage IV disease at initial diagnosis, but had interval development of stage IV (distant metastatic) disease prior to the age (time) of specimen.*

* + - IF SpecimenSiteOfOrigin contains [“eye” OR “choroid” OR “ciliary body” OR “conjunctiva”] AND {PathGroupStage does *not* contain “IV” OR ClinGroupStage does *not* contain “IV”} AND [{combined entry on Metastatic file with: MetsDzPrimaryDiagnosisSite contains “eye” OR “choroid” OR “ciliary body” OR “conjunctiva”] AND MetastaticDiseaseInd = [“Yes – Regional” OR “Yes – NOS”] AND [AgeAtMetastaticSite <= AgeAtSpecimenCollection OR AgeAtMetastaticSite = “Age Unknown/Not Recorded”]}, THEN **AssignedStage** = “III”

*Example: 8OR7RX5NO5*

*NOTES: This rule defines patients who did not have stage IV disease at initial diagnosis, and have had nodal recurrence of ocular melanoma (rare).*

* + **Define the patients with primary mucosal melanoma** 
    - If PrimaryDiagnosisSite contains “sinus” OR “gum” OR “nasal” OR “urethra” then **AssignedPrimarySite** = mucosal
      * The list only includes the primary mucosal sites present in this dataset and not all possible mucosal sites for melanoma in general.
    - **Assign the stage for these patients using the following rules. All of these rules assume AssignedPrimarySite = mucosal** 
      * IF PathGroupStage contains “IV”, then **AssignedStage** = “IV”
      * IF ClinGroupStage contains “IV” AND PathGroupStage is [“Unknown/Not Reported” OR “No TNM applicable for this site/histology combination” OR “Unknown/Not Applicable”], THEN **AssignedStage** = “IV”

*NOTES: The above two rules define the patients with stage IV disease at initial diagnosis, so the specimen is stage IV regardless of the time interval between diagnosis and specimen collection.*

* + - * IF [PathGroupStage does *not* contain “IV” OR ClinGroupStage does *not* contain “IV”] AND Primary/Metastatic = “Primary”, THEN **AssignedStage** = *numerical* value of PathGroupStage OR ClinGroupStage (if PathGroupStage is [“Unknown/Not Reported” OR “No TNM applicable for this site/histology combination” OR “Unknown/Not Applicable”])

*Examples: EKZGU61JTP, W15ISRT8YI, WLYUAJ9AAE*

*NOTES: This rule defines patients without stage IV disease who have a tumor specimen from the primary mucosal site at diagnosis.*

* + - * IF [PathGroupStage does *not* contain “IV” OR ClinGroupStage does *not* contain “IV”] AND Primary/Metastatic = “Metastatic” AND {combined entry on Metastatic file with: MetastaticDiseaseInd = “Yes – Distant” AND [AgeAtMetastaticSite <= AgeAtSpecimenCollection OR AgeAtMetastaticSite = “Age Unknown/Not Recorded”]}, THEN **AssignedStage** = “IV”

*Example: Z7CEUA8SAJ*

*NOTES: This rule defines patients without stage IV disease who have a tumor specimen obtained after interval development of stage IV disease.*

* + **Define the patients with primary melanoma of unknown origin** 
    - If PrimaryDiagnosisSite contains ‘unknown’, then **AssignedPrimarySite** = unknown
    - **Assign the stage for these patients using the following rules. All of these rules assume AssignedPrimarySite = unknown. Note that all specimens from unknown primaries in this dataset are from stage IV disease.** 
      * IF PathGroupStage contains “IV”, then **AssignedStage** = “IV”
      * IF ClinGroupStage contains “IV” AND PathGroupStage is [“Unknown/Not Reported” OR “No TNM applicable for this site/histology combination” OR “Unknown/Not Applicable”], THEN **AssignedStage** = “IV”
  + **Once AssignedPrimarySite and AssignedStage have been populated for all patients in Group A, then can use Age At Specimen Collection to determine ICB status of tumor and patient as previously done.**

1. **For Group B**: 1 melanoma diagnosis and >1 tumor sequenced
   * **First, determine which sequenced tumor to use:** 
     + If RNAseq data is missing for all tumors for a patient, then exclude from analysis.
     + If RNAseq is available for just one tumor, select the tumor with RNAseq data
     + If RNAseq data is available for > 1 tumors:
       - If the patient has a tumor with SpecimenSiteofCollection that contains “skin” AND does not also have a tumor with SpecimenSiteofCollection that contains either “lymph node” or “soft tissue”, then select the tumor with SpecimenSiteofCollection that contains “skin”
       - If none of the patient’s tumors have SpecimenSiteofCollection that contains “skin” or “soft tissue”, then select the tumor with SpecimenSiteofCollection that contains “lymph node”
       - If a patient has a tumor with a SpecimenSiteofCollection that contains [“skin” OR “soft tissue”] AND a tumor with a SpecimenSiteofCollection that contains “lymph node”, then select the one with earliest Age At Specimen Collection
       - If the patient does NOT have any tumor with a SpecimenSiteofCollection that contains [“skin” OR “soft tissue” or “lymph node”], then select the tumor with the earliest Age At Specimen Collection
   * **Next, determine AssignedPrimarySite and AssignedStage for the selected tumor and diagnosis pairing using the same rules as listed under Group A**
2. **For Group C:** >1 melanoma diagnosis and 1 tumor sequenced
   * **First, determine which diagnosis to use:** 
     + For specimens with the field Primary/Met = “Primary”
       - IF AgeAtSpecimenCollection is WITHIN 90 days AFTER the AgeAtDiagnosis for only one of the diagnoses, then use that diagnosis.
         * *Examples: 2FEX3JHKCW, 8BCBHGEO4N, FNP53S81KA*
       - IF AgeAtSpecimenCollection is WITHIN 90 days AFTER the AgeAtDiagnosis for more than one of the diagnoses, but the PrimaryDiagnosisSite is NOT the same for all diagnoses, then use the diagnosis that has a PrimaryDiagnosisSite = SpecimenSiteOfCollection.
         * *Example: ILE2DL0KMW*
       - IF AgeAtSpecimenCollection is GREATER THAN 90 days AFTER the AgeAtDiagnosis for all of the diagnoses OR if the AgeAtDiagnosis for one or more of the diagnoses is unknown, but the Histology is not the same for all diagnoses, then use the diagnosis that has a Histology = specimen Histology/Behavior (non-numerical portion of that field)
         * *Example to confirm with Slingluff: L2R9RJJ88C*
     + For specimens with the filed Primary/Met = “Metastatic”
       - IF SpecimenSiteOfCollection does not contain [“soft tissues” OR “lymph node”], then use the diagnosis with the earliest AgeAtDiagnosis.
         * *Examples: 087FO3NF65, AC2EJBKWJO*
         * *Examples to confirm with Slingluff: 9NOLH4M870, HLIXS3VDZ6, SHTJKKY76C*
       - IF SpecimenSiteOfCollection contains “soft tissues” and AgeAtSpecimenCollection is WITHIN 90 days AFTER the AgeAtDiagnosis, then use the diagnosis WITHIN 90 days AFTER the AgeAtDiagnosis.
         * *Example: DEB9M36STN*
       - IF SpecimenSiteOfCollection contains “lymph node” AND AgeAtSpecimenCollection is WITHIN 90 days AFTER the AgeAtDiagnosis for any of the diagnoses AND only one diagnosis has PathNStage that does NOT contain [“N0”, “Nx” OR “Unknown/Not Applicable” OR “No TNM applicable for this site/histology combination”], then use that diagnosis (the one with known positive nodes on PathNStage).
         * *Example: 5BS8L7PCCE, 39TYSJBNKK, 6RX3G5GV02, 643X8OLYWR, 7HU06PZK4Q*
       - IF SpecimenSiteOfCollection contains “lymph node” AND AgeAtSpecimenCollection is GREATER THAN 90 days AFTER the AgeAtDiagnosis for all of the diagnoses OR if the AgeAtDiagnosis for one or more of the diagnoses is unknown AND the PrimaryDiagnosisSite for only one of the diagnosis = MetsDzPrimaryDiagnosisSite, then use the diagnosis associated with that PrimaryDiagnosisSite
         * *Examples: 383CIRVHH2, KWLPMWV0FM*
   * **Next, determine AssignedPrimarySite and AssignedStage for the selected diagnosis and tumor pairing using the same rules as listed under Group A**
3. **Group D**: >1 melanoma diagnosis and >1 tumor sequenced
   * Start with the rules outlined in #3 (for Group B) to select the tumor of interest.
   * Then, use the rules outlined in #4 (for Group C) to select the diagnosis of interest to pair with the tumor that has been selected.
   * Next, determine **AssignedPrimarySite** and **AssignedStage** for the selected tumor and diagnosis pairing using the same rules as listed under Group A