***Título:*** *“Predicción del riesgo individual de recaída en pacientes con cáncer colorectal localmente avanzado tratados con quimioterapia neoadyuvante y cirugía tras mediante la aplicación de algoritmos de inteligencia artificial*”.

1. ***ANTECEDENTES Y ESTADO ACTUAL***

El CCR supone un problema epidemiológico y sanitario importante dada su incidencia (segundo tumor tras el cáncer de mama en mujeres y el de próstata en varones) y mortalidad (segundo tras el de pulmón). Alrededor del 20% de los pacientes presentan MH en el momento del diagnóstico y más del 50% las desarrollan a lo largo de la evolución de su enfermedad (15-25% de forma sincrónica y 50-60% metacrónica). Además, las complicaciones derivadas de las MH constituyen la principal causa de muerte de estos enfermos.

Aunque la cirugía es el único tratamiento potencialmente curativo en estadios I-III, existe un amplio consenso respecto al empleo de quimioterapia adyuvante con esquemas basados en una combinación de oxaliplatino y fluoropirimidinas. Sin embargo, en los últimos años se ha extendido el empleo de quimioterapia preoperatoria o neoadyuvante (QTNA), ya que se asocia con una significativa reducción en la afectación ganglionar, un incremento en las tasas de R0 y una tendencia a un menor riesgo de recaída respecto a un enfoque convencional. Tan sólo la inestabilidad de microsatélites y el grado de regresión histológica tras la QTNA se han correlacionado hasta la fecha con el pronóstico. Identificar subgrupos de pacientes que puedan beneficiarse en mayor medida de esta estrategia es, a día de hoy, una prioridad. Para ello en el presente trabajo planteamos incorporar datos farmacogenéticos, parámetros cinéticos, análisis de microbioma intratumoral, parámetros radiómicos y algoritmos de machine learning a una serie de pacientes con CCLA tratados mediante QTNA.

***1-Farmacogenética***. En los últimos años se han caracterizado mediante análisis trancriptómicos diferentes subgrupos moleculares (CMS) dentro de CCR metastásico que tienen implicaciones predictivas y pronósticas. El correlato inmunohistoquímico de estos análisis transcriptómicos ha sido recientemente publicado, sugiriéndose que mediante la determinación IHQ de la expresión de 5 proteínas se puede clasificar correctamente a los pacientes en los CMS con una probabilidad de acierto del 80% El papel pronostico y predictivo de este panel inmunohistoquímico y de los subgrupos moleculares subyacentes aún no ha sido analizado en el escenario clínico que nos ocupa.

***2-Farmacocinética***. En pacientes tratados con FOLFOX, la dosificación del 5FU se realiza a través del denominado método tradicional de la superficie corporal (BSA), que considera tan sólo peso y talla de cada paciente. El 5-Fluorouracilo presenta una amplia variabilidad intra e interindividual, existiendo una clara relación entre sus niveles plasmáticos, su toxicidad y su eficacia. La dosificación basada en parámetros farmacocinéticos calculada para alcanzar una determinada área bajo la curva plasmática ha demostrado incrementar la eficacia y reducir el perfil de toxicidad, tanto en pacientes con CCR metastásico como en aquellos con CCLA tratados con cirugía y FOLFOX adyuvante. El papel de la optimización posológica del 5-FU dentro de un programa de QTNA en pacientes con CCLA está hoy por definir y podría ser prometedor.

***3-Análisis in silico*** mediante algoritmos de machine learning para generar un modelo poblacional de predicción individual de recaída en pacientes con metástasis hepáticas de CCR resecadas tras tratamiento neoadyuvante. Los algoritmos de aprendizaje supervisado generan conocimiento a partir de datos en los que conocemos la variable respuesta a predecir. A partir de una base de datos, donde se le facilita diversos ejemplos al modelo junto con la variable respuesta conocida, los algoritmos de aprendizaje supervisado generan modelos de predicción. Estos modelos generan conocimiento aprendiendo de los datos (proceso que suele llamarse entrenamiento del algoritmo). Una vez el modelo está entrenado, este nos permite conocer la variable respuesta de nuevos pacientes, que no han sido empleados en el entrenamiento, dicho de otra forma, nos permite clasificarlos o hacer una predicción individualizada sobre ellos. El desarrollo de herramientas de predicción de riesgo individual de recaída precoz podría ayudar a optimizar y guiar la toma de decisiones terapéuticas. Hasta la fecha, pocos estudios oncológicos basados en el uso de estos datos han sido publicados. Sin embargo, esta técnica podría ser una alternativa potencial a los métodos de regresión multivariable que habitualmente predominan en la literatura científica oncológica

**4-Análisis del microbioma.** El microbioma en humanos contiene trillones de organismos distribuidos por todo el cuerpo, que pueden ser caracterizados con técnicas de secuenciación masiva (16S ribosomal *RNA* *sequence*, *shotgun sequencing*, metatranscriptómica, metaproteinómica, etc.), lo que constituye una aproximación a la composición relativa de una determinada comunidad bacteriana. En los últimos años se ha descrito un claro vínculo entre el microbioma y el cáncer que afecta a varios aspectos relacionados con la oncogénesis, eficacia de los fármacos antitumorales y pronóstico. Los microorganismos que componen la microbiota pueden encontrarse dentro del propio tumor, en el tejido sano adyacente o en el órgano donde se asienta (intestino, vía área, etc.). Aunque en el contexto del CCR el aspecto más estudiado es la microbiota intestinal o fecal, existe evidencia del papel que desempeña la microbiota intratumoral en esta enfermedad. El microambiente tumoral es hipóxico, rico en áreas necróticas capaces de generar gradientes quimiotácticos para algunos microorganismos, presenta vasculatura aberrante e hiperpermeable y un ambiente predominantemente inmunosupresor. Todos estos factores favorecen el paso de microorganismos al interior de los tumores y su crecimiento.

La correlación entre la composición de la microbiota fecal e intratumoral no es del todo coincidente, con diferencias descritas en la presencia de *Fusobacterium*, *Providencia* y *Acinetobacter*, entre otros. Estas bacterias intratumorales en el CCR se asocian con el grado de diferenciación tumoral, el fenotipo CIMP (*CpG island methylator phenotype*), el grado de infiltración por el sistema inmune, así como con la eficacia y tolerancia a la quimioterapia. Además, se ha demostrado que en el proceso en el que el CCR metastatiza al hígado, se “transporta” parte de su microbioma a través de las células tumorales circulantes. El efecto que tiene la interacción entre el microbioma y las células tumorales presentes en las MH está hoy en día poco estudiado. En modelos preclínicos, las bacterias intratumorales pueden modular la inmunidad innata dentro del microambiente tumoral mediante los PRRs (*pattern recognition receptors*), lo que impacta sobre la sensibilidad a algunos fármacos de quimioterapia. En modelos preclínicos de CCR, *F. nucleatum* modifica la respuesta a 5-fluorouracilo y oxaliplatino interaccionando con la señalización de TLR4 y MYD88, así como con diversos miRNAs (probablemente mediante la activación de la cascada de la autofagia). En un escenario clínico, el análisis de la microbiota obtenida a partir de muestras de parafina de 85 pacientes con cáncer de colon localizado demostró que la presencia de una OTU (unidad taxonómica) enriquecida en clostridiales se correlacionó con una menor infiltración tumoral por linfocitos CD8+ y, por tanto, con un incremento en el riesgo de recidiva.

**5- Radiómica *(radiomics).*** Las técnicas de imagen empleadas en oncología (la resonancia magnética nuclear, la tomografía computarizada o la tomografía por emisión de positrones) proporcionan información crucial para representar con precisión los tumores y su microambiente de una manera no invasiva. En general, se evalúan diferentes características de imagen, como el tamaño, bordes e irregularidades del tumor. Sin embargo, esta evaluación crea descripciones subjetivas de la enfermedad y por ello, existe una considerable variabilidad interobservador. Como consecuencia de la necesidad de reflejar objetivamente las diferentes características del tumor y su microambiente surge la radiómica, que tiene como objetivo extraer características de imágenes de alto rendimiento para una definición tisular completa y automatizada. El empleo de esta técnica posibilita la transformación de datos encriptados de la imagen en información valiosa que posteriormente puede ser utilizada para predecir la respuesta del tumor a un determinado tratamiento, así como el riesgo de recaída de cada paciente tras una maniobra terapéutica concreta.

***2. HIPÓTESIS Y OBJETIVOS DEL PROYECTO***

***Hipótesis***

1. El manejo convencional del CCLA consiste en cirugía seguido de quimioterapia adyuvante, aunque el empleo de quimioterapia preoperatoria se correlaciona, en algunos escenarios clínicos, con una menor morbilidad quirúrgica y un menor riesgo de recaída a dos años. Analizaremos de manera retrospectiva la experiencia de nuestro centro con este último enfoque, en términos de factibilidad, perfil de efectos secundarios, morbimortalidad, tasa de R0, grado de regresión histopatológica, supervivencia libre de progresión, supervivencia global y patrones de recaída tras un seguimiento prolongado
2. La composición individual de la microbiota intratumoral presente en pacientes con CCR, a través de su efecto modulador sobre el sistema inmune, podría ser en parte responsable del riesgo diferencial de recaída.
3. El análisis de parámetros radiómicos, obtenidos a partir de las tomografías seriadas realizadas durante el tratamiento neoadyuvante, mediante técnicas de procesamiento de imagen con inteligencia artificial (IA) podría generar información relevante que pudiera ser integrada en algoritmos de predicción de riesgo.
4. La creación de un modelo poblacional de predicción del riesgo individual de recaída en pacientes con CCR a través de la IA mediante la incorporación de variables clínicas, analíticas, parámetros de microbioma y empleo de radiómica podría ayudar a optimizar los abordajes terapéuticos actuales.

***Objetivos***

1. Caracterizar mediante análisis inmunohistoquímico de biopsias pre-tratamiento obtenidas mediante colonoscopias diagnósticas los diferentes subgrupos moleculares descritos en CCR y definir su papel pronóstico y/o predictivo en el escenario neoadyuvante.
2. Calcular parámetros farmacocinéticos del 5-Fluorouracilo administrado dentro del esquema FOLFOX preoperatorio en pacientes con CCLA e identificar un AUC target que se correlacione con un beneficio en términos de eficacia y toxicidad y que pueda ser empleado de manera prospectiva en futuros estudios.
3. Localizar en las muestras tumorales obtenidas tras la resección de las MH, bacterias y hongos mediante hibridación in situ con secuencias únicas capaces de identificar estos microorganismos empleando el análisis del perfil taxonómico con amplicones 16S (bacterias) o ITS (hongos). Analizar las diferencias en los microbiomas de los largos y cortos supervivientes.
4. Describir durante el tratamiento neoadyuvante, la dinámica tumoral mediante parámetros de radiómica y su correlación con el riesgo de recaída.
5. Desarrollar un modelo poblacional de predicción individual de recaída, mediante la aplicación de algoritmos de *machine learning*, con el propósito de integrar todos los parámetros previamente estudiados para definir grupos de pacientes con diferente comportamiento clínico.

***Plan de Trabajo y Metodología***

1. **Análisis retrospectivo de la serie clínica**: Tras la elaboración de una base de datos extraída de los pacientes tratados en nuestro centro hospitalario en el servicio de oncológica médica, se determinarán una serie de variables que nos permitan llevar a cabo nuestro proyecto. Las variables consideradas en cada uno de los pacientes incluídos, han sido sexo, edad, ECOG, superficie corporal, fecha del diagnóstico, localización del tumor primario, estadio clínico (cTNM) fármacos empleados, fecha de inicio y fin, número de ciclos administrados, dosis empleadas, intensidad de dosis (incluyendo intensidad de dosis ideal, intensidad de dosis real y el cociente de ambas, complicaciones secundarias que precisaran ingreso (si/no), uso de factores estimulantes de colonias granulocíticas (si/no), niveles de pirimidinas endógenas, área bajo la curva de 5-fluorouracilo, tipo de cirugía, estadio patológico (pTNM), grado de regresión tumoral (MSKCC), grado de regresión tumoral ganglionar (adaptada del sistema de gradación de Miller & Payne), número de ganglios resecados, número de ganglios afectados, ratio de afectación ganglionar (LNR), ratio de afectación ganglionar modificado (LNR modificado) que corresponde con la fórmula 1 + nº ganglios afectados / 1 + nº ganglios resecados, invasión vascular (si/no), invasión perineural (si/no), márgenes quirúrgicos (R0/R1), adyuvancia (si/no), tipo de adyuvancia, analíticas evolutivas obtenidas durante el tratamiento y el seguimiento clínico, progresión en los dos primeros años tras la cirugía (si/no), fecha de progresión, tipo de progresión (local y/o a distancia).
2. **Inmunohistoquimica**: Para el estudio inmunohistoquímico de los distintos subgrupos moleculares se seleccionarán del archivo del Servicio de Anatomía Patológica, bloques de parafina con presencia de tumor viable. Se realizarán secciones de 4μm de espesor, con el respectivo pre-tratamiento de las laminillas (desparafinización y recuperación antigénica). Los anticuerpos primarios a estudiar, aplicando sus respectivos períodos de incubación, serán los siguientes: anti-CDX2 (clon EPR2764Y; No requiere dilución; Ventana Ref. 760-4380), anti-FRMD6 (1:500; Sigma; HPA001297), anti-HTR2B (1:75; Sigma; HPA12867), anti-ZEB1 (1:500; Sigma; HPA027524) y anti-Pan Keratin (AE1/AE3; Roche; PCK26KER). Posteriormente, se aplicará el anticuerpo secundario+cromógeno (sistema Ventana ultraView Universal DAB Detection Kit), con un segundo período de incubación y finalmente se realizará la contratinción con hematoxilina. Las preparaciones serán evaluadas por un patólogo, estimando la intensidad de inmunotinción, siguiendo la clasificación Histocore: 0:negativo; 1:débil; 3:moderado; 4:intensidad fuerte. Además, se estimará el porcentaje de células positivas (0-100%), en intervalos de 10%.
3. **Farmacocinética**: La extracción de las muestras biológicas (sangre) se realiza durante la administración del 5-fluorouracilo mediante venopunción en el brazo contralateral al de la administración del fármaco. Se realizan tres extracciones diferentes en los dos primeros ciclos. Una vez extraídas, manteniendo la cadena de frío, serán trasladadas a la unidad de farmacocinética para su centrifugación. Esta centrifugación se realiza a 4ºC y 2.500 revoluciones por minuto (r.p.m.), durante 10 minutos. Posteriormente son almacenadas a -30ºC para su posterior análisis. Las concentraciones plasmáticas del 5-Fu se miden por cromatografía líquida de alta resolución (HLPC) de acuerdo con la técnica validada en nuestro centro. Tras acidificar la muestra con 20 microlitros de ácido ortofosfórico al 5%, el fármaco se extrae del plasma mediante n-propanol-dietil éter. La fase orgánica se evapora por sequedad y el residuo se disuelve en 100 microlitros de fase móvil (formada por dihidrogenofosfato de potasio). Finalmente, 20 microlitros de la mezcla obtenida son inyectados en la columna analítica del dispositivo de cuantificación. La optimización de los datos de concentración plasmática – tiempo se realiza por metodología bayesiana, utilizando un modelo poblacional monocompartimental lineal desarrollado por la Unidad de Farmacocinética Clínica de nuestro Centro en el año 2002 e implantado en el paquete de programas USC\*PACKv11.2 de la Universidad del Sur de California. Este modelo fue desarrollado y validado en pacientes con cáncer colorrectal y biliopancreático. Los datos de concentración plasmática de 5-Fluorouracilo se ajustaron tanto a un modelo farmacocinético lineal como no lineal, no encontrándose diferencias significativas en los criterios de selección de los modelos. En nuestro caso, dado que la administración del 5-Fluorouracilo dentro del esquema FOLFOX se realiza mediante perfusión continua de 46 horas, se ha utilizado un modelo monocompartimental lineal para el cálculo individual de los parámetros farmacocinéticos. Durante el tratamiento administrado, se determinan al menos los parámetros farmacocinéticos del primer y segundo ciclo de tratamiento. El motivo de realizar una segunda determinación es evaluar la variabilidad intercíclica y confirmar la estimación realizada. En caso de identificarse diferencias significativas entre los parámetros estimados en el primer y el segundo ciclo que no son justificables por los datos clínicos referidos por el paciente, se procede a la monitorización de un tercer ciclo o posteriores.
4. **Estudio computacional**: La metodología que se emplea comienza por la recogida de datos de novo. Con el objetivo de evitar que el proceso de aprendizaje de los algoritmos empleados se focalice excesivamente en determinadas características particulares de las variables de entrenamiento recogidas (concepto conocido como *“overfitting”*), solo serán evaluadas aquellas consideradas como más relevantes en relación con el riesgo individual de recidiva después de la cirugía. Para que los algoritmos de *machine learning* sean capaces de predecir un desenlace, la relación entre el número de variables estudiadas y el número de individuos analizados debe estar entre 1/5 y 1/10. En un segundo tiempo, sobre estas variables seleccionadas inicialmente se llevará a cabo un análisis exploratorio univariante y, finalmente, se elegirán las más relevantes para entrenar cada uno de los diferentes algoritmos de *machine learning* utilizados. Para el desarrollo del modelo poblacional de predicción individual de recaída en MH de CCR resecadas tras tratamiento neoadyuvante, se emplearán diferentes algoritmos de *machine learning*: regresión logística, árbol de decisión, *random forest*, máquina de vector soporte y *k*-vecinos cercanos. El *software* empleado para ensayar los diferentes algoritmos será *R-Project* con la interfaz de uso *R-Studio*, aplicando diferentes librerías en función de los algoritmos empleados (*caret*, *car*, *rpart,* *random forest*, *e1071*). Además, para la visualización de las curvas ROC y ABC-ROC se empleará la librería pROC. Los modelos predictivos basados en los algoritmos de regresión logística, árbol de decisión y máquina de vector soporte serán validados con la técnica *k-Fold Cross-Validation* (en nuestro caso *5-Fold Cross-Validation*). El modelo predictivo basado en el algoritmo de *random forest* será validado con la técnica de *bagging* y el relacionado con el algoritmo de *k*-vecinos cercanos será con la técnica de *lazy-learning*. Estas técnicas de validación han sido utilizadas en diferentes trabajos publicados hasta la fecha dentro del campo de la medicina. De manera adicional, se plantea llevar a cabo una validación externa del modelo poblacional definitivo, empleando el algoritmo con mayor capacidad predictiva de todos los estudiados. Se considera que el número necesario para llevar a cabo la validación externa debería corresponder con el 20% del total de la cohorte estudiada, dado que este es el porcentaje considerado correcto para la realización de validaciones externas de algoritmos de *machine learning*. El aprendizaje de los algoritmos de *machine learning*, va a depender de la cantidad y calidad de los datos a partir de los que aprenda. Aunque la relación entre el aprendizaje y la cantidad de los datos es directamente proporcional, no es una relación lineal simple con una pendiente constante. La forma de la curva tendrá una pendiente muy pronunciada al principio y posteriormente se convertirá en una meseta, donde la pendiente será próxima a cero y no merecerá la pena seguir aumentando el tamaño de la muestra. Se puede emplear un modelo de regresión para conocer o estimar el tamaño muestral idóneo para llegar a ese punto donde no es necesario seguir aumentando la muestra. Este modelo puede construirse a partir de los valores de *accuracy* (variable a predecir) obtenidos a partir de distintos tamaños muestrales disponibles. Empleando esta técnica, estimamos que el tamaño muestral necesario para llegar al *accuracy* del 80% está en torno a unos 90-100 pacientes con un intervalo de confianza del 95%. 5-**5**-**Análisis del microbioma tumoral**

Se llevarán a cabo los siguientes pasos: 1- Identificación y registro de muestras

en el laboratorio de Anatomía Patológica. 2- Extracción de ADN y control de calidad, usando kits específicos para cada tipo de muestra. La extracción de ADN de muestras de FFPE (*formalin-fixed paraffin-embedded*) se pondrá a punto usando varios protocolos. Se evaluará con determinaciones de concentración con Qubit® y electroforesis en gel de agarosa. 3- Preparación de las librerías y secuenciación. Se procederá a la identificación de los microorganismos presentes con el análisis de secuencias específicas para hongos (secuencias ITS) y para bacterias (16S) mediante la amplificación de las secuencias correspondientes con biomarcadores de interés, el etiquetado de las moléculas que correspondientes a cada muestra y la adición de los adaptadores necesarios para la secuenciación. Este proceso permite la identificación de distintos grupos taxonómicos con la posibilidad de alcanzar niveles resolutivos de cepa. Para ello, se procede al estudio por PCR (reacción en cadena de la polimerasa) de las secuencias con biomarcadores de hongos y bacterias etiquetando los productos de amplificación con una secuencia *tag* específica que permitirá examinar múltiples muestras de forma simultánea y obtener finalmente el resultado por separado de cada determinación. La secuenciación de tales biomarcadores se puede llevar a cabo mediante tecnología de *Illumina*, que proporciona un mayor rendimiento y menor coste, centrándose en determinadas regiones hipervariables que ofrecen una capacidad de resolución útil.

**6-Radiómica.** Para la creación del algorítmo de inteligencia artificial (IA) y radiómica *(“deep radiomics”)*, se utilizarán imágenes de las tomografías seriadas durante la neoadyuvancia (basal y previa a la cirugía hepática). Para la correcta extracción de las características radiómicas de las imagénes de la tomografía y su posterior análisis con algoritmos de IA, la plataforma final a diseñar constará de diferentes etapas;

Segmentación de la región de interés: en este paso, la problemática se basa en la identificación de variaciones en las características radiómicas en relación con las áreas tumorales y su microentorno. Para ello, un radiólogo experto se encargará de revisar imágenes y delimitar las regiones de referencia que se analizarán posteriormente. Estas segmentaciones tendrán una dimensión mínima determinada para la correcta extracción de las características radiómicas en todas las imágenes.

Para que el algoritmo a diseñar sea lo más robusto posible y así predecir con exactitud los subgrupos riesgo de reaparición de la enfermedad, es necesario realizar un filtrado de características y eliminar las variables menos relevantes. Para ello, se realizarán estudio de correlación y técnicas de *machine learning* para ver el impacto que tiene cada rasgo en particular respecto al problema planteado.

Diseño e implementación del algoritmo de radiómica: una vez se han seleccionado los biomarcadores de mayor impacto, estos se unirán mediante una matriz a las imágenes de segmentación de la región de interés de inicio y servirán como punto de partida para el aprendizaje del modelo final de radiómica. En el caso de que el número de imágenes sea limitado, se implementará la técnica de aumentación de datos *(data augmentation)* y se adoptará la validación cruzada para disminuir los efectos de “sobreentrenamiento” de la red.

***Plan de trabajo***

Adjuntamos el plan de trabajo planificado con una duración aproximada de 2 años. Se divide el tiempo estimado en función de la complejidad y los requerimientos de cada parte a desarrollar, tras la valoración conjunta por los integrantes del equipo investigador. Todos los miembros del proyecto tendrán una implicación directa y trabajarán de forma coordinada en cada fase establecida:

* **Meses 1-9:**
* Elaboración de la base de datos considerando todas las variables clínicas, analíticas, histopatológicas y cinéticas.
* Estudio del microbioma tumoral de las metastasis hepáticas resecadas.
* Análisis de parámetros radiómicos extraídos de las tomografías seriadas realizadas durante el periodo de quimioterapia neoadyuvante.
* **Meses 10-16:** Generación de diferentes modelos de predicción y entrenamiento con los diversos algoritmos de machine learning.
* **Meses 17-19:** Elección del modelo con mayor capacidad predictiva.
* **Meses 20-22:** Validación del modelo elegido con una serie adicional de pacientes con las mismas características que la original.
* **Meses 23-24:** Publicación de los resultados.

***Medios disponibles para la realización del proyecto***

* **Experiencia del investigador y del centro sobre el tema:** Este proyecto cuenta con la colaboración entre la Clínica Universidad de Navarra (CUN), Centro de Investigación Médica Aplicada (CIMA) y Universidad de Navarra (UNAV). Los principales departamentos implicados son: Radiología, Farmacia y Nutrición, Anatomía Patológica, Cirugía General, y especialmente el departamento de Oncología Médica. Se dispone de la colaboración de *NNBi S.L*. (unidad de I+D) y como soporte de la parte de ingeniería biomédica para la incorporación de algoritmos de IA.
* **Equipos para extracción de DNA, preparación de librerías y secuenciación:** CIMA LAB *Diagnostics* dispone de recursos y equipamiento que van desde la tecnología genética convencional (pirosecuenciadores, secuenciadores capilares, PCR estándar y tiempo real), hasta los equipos más novedosos para el análisis genómico masivo. En la actualidad se dispone de plataformas de genómica complementarias, tanto de *Illumina (MiSeq* y *NextSeq* *500)* como de *Thermo Fisher* *(Ion S5™ System)*. Cuenta además con capacidad bioinformática de almacenamiento y análisis y dispone del *software* necesario para el análisis de experimentos de Next-Generation Sequencing, tanto paneles como secuenciación de exomas o *RNA-Seq*. Los colaboradores externos cuentan con la experiencia y las instalaciones necesarias para llevar a cabo sus respectivos estudios. Asimismo, el departamento de Anatomía Patológica aporta la ayuda requerida para dirigir y gestionar las muestras.
* **Equipos para estudios farmacocinéticos:** El equipo cromatografía líquida de alta resolución *(HPLC),* modelo HP 1100 con un detector de matriz de diodos.
* **Soporte y gestión de datos:** se utilizan diferentes programas (Microsoft Excel, IBM SPSS® *Statistics* y Stata). Se dispone de ayuda informática para la trasmisión codificada de datos desde el sistema CUN (programa informático del centro).

***Justificación detallada de la ayuda solicitada***

Para llevar a cabo nuestro estudio, se dividen los recursos necesarios propuestos para el desarrollo de una herramienta de predicción individual del riesgo de recidiva en los siguientes apartados:

1. Estudio de **microbioma tumoral** en las muestras tumorales. Este proceso, se resumen en las siguientes etapas:

* Selección, identificación y registro.
* Extracción del ADN y realización de controles de calidad.
* Elaboración de las librerías:
  + Identificación de los microorganismos
  + Indexado
  + Demultiplexado
* Secuenciación de *Illumina*

1. Creación de un algorítmo de IA y **radiómica** con el propósito de una adecuada exportación de las imágenes y características radiómicas, se sintetiza el proceso en varias fases:

* Segmentación de la región de interés.
* Preprocesado para conseguir datos sean homogéneos y facilitar la creación de un algoritmo radiómico óptimo.
* Extracción y selección de características radiómicas:
  + *“First order features”* o características de histograma.
  + *“Second order features”* o características texturales.
  + *“High order features”* para alcanzar patrones sofisticados.que se emplean para la consecución de patrones sofisticados de la imagen.
* Filtrado y optimización de datos con la ayuda de la técnica de *machine learning*.
* Diseño e implementación del algorítmo de radiómica.
* Integración de resultados para identificar subgrupos de riesgo de recaída en CCR y MH resecadas.

1. El proceso del *machine learning* consta de forma resumida de los siguientes pasos:

* **Preprocesamiento** de la base de datos: adaptación del registro a las necesidades del análisis, tratamiento de datos perdidos, inconsistencias, redundancias, creación de variables derivadas, etc.
* **Análisis descriptivo** de la base de datos según el tipo de cada una de las covariables:
  + Discretas: frecuencias, porcentajes relativos, diagramas de barras, etc.
  + Continuas: estimadores de tendencia central, dispersión, etc.
  + Este análisis se realizará a nivel: descriptivo, inferencial y visual.
* **Definición del objetivo** a predecir y las características del estudio de *machine learning*.
  + Definir objetivo y posibles variables predictoras.
  + Decidir técnicas a implementar y de validación (dependerán del tamaño muestral).
* **Implementación de los algoritmos** de *machine learning* y las pruebas de validación definidas en *Python* y/o *R*.
* **Valoración de los resultados** obtenidos de los modelos anteriormente implementados y estimación de los mejores modelos a través de los resultados de estos (*accuracy*, AUC-ROC, especifidad, sensibilidad, etc.).

***Bibliografía***

Siegel RL, Miller K, Goding A, et al. Colorectal Cancer Statistics, 2020. CA Cancer J Clin 2020; 0: 1–20.

NCCN Clinical Practice Guidelines in Oncology. Colon Cancer. Version 3, 2020. May 6, 2020.

Picardo S, Coburn B, Hansen A, et al. The microbiome adn cancer for clinicians. Critical Reviews in Oncology/Hematology 2019; 141: 1-12.

Helmink BA, Wadud Khan MA, Hermann A, et al. The microbiome, cancer, and the cancer therapy. Nature Medicine 2019; 377-388.

Xavier JB, Young VB, Skufca J, et al. The cancer microbiome: Distinguishing direct and indirect effects requires a systematic view. Trends in cancer, Cell 2020; Mar; 6: 192-204.

Cogdill AP, Gaudreau PO, Arora Reetakshi, et al . The impact of intratumoral and gastrointestinal microbioma on systemic cancer therapy. Trends in Immunology 2018; 39: 900-920.

Bruno R, Bottino D, P de Alwis D, et al. Progress and opportunities to advance clinical cancer therapeutics using tumor dynamic models. Clin Cancer Res 2020; 26: 1787-95.

Michaelis LC, Ratain MJ. Measuring response in a post-RECIST world: from black and white to shades of grey. Nat Rev Cancer 2006; 6: 409–14.

Reiter MJ, Hannemann NP, Schwope RB, et al. Role of imaging for patients with colorectal hepatic metastases: what the radiologist needs to know. Abdom Imaging 2015; 40: 3029–42.

Brennan M, Kattan M, Klimstra D et al. Prognostic nomogram for patients undergoing resection for adenocarcinoma of the pancreas. Annals of Surgery 2004; 2: 293-298.

Hayward J, Álvarez S, Ruiz C et al. Machine learning of clinical of clinical performance in a pancretic cancer database. Artificial Intelligence in Medicine 2010; [Volume 49, Issue 3](https://dl.acm.org/toc/aiim/2010/49/3).

Sala P, Oyaga-Iriarte E, Yu KH, et al. Use of Machine-Learning Algorithms in intensified preoperative therapy of pancreatic cancer to predict individual risk of relapse. [Cancers (Basel)](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6562932/); 2019; 30; 11: 606.