

IA4DT2



IA4DT2

Para poder hacer realidad este proyecto, cuyo objetivo es generar modelos de predicción de Diabetes de Tipo 2 (DT2), combinando la información genética, de estilo de vida, ambiental y clínica, se ha establecido una colaboración entre el grupo de Genética Computacional del Barcelona Supercomputing Center (BSC), el grupo GCAT|Genomes for Life del Institut Germans Trias i Pujol (IGTP) y el departamento de Derecho Público de la Universidad del País Vasco (UPV/EHU).

La generación de modelos de predicción de DT2 precisa de una gran cantidad de datos genéticos, clínicos y ambientales. De hecho, para asegurar que los resultados de este estudio sean reproducibles y aplicables a la población en general, es necesario disponer de la misma información de pacientes en dos cohortes independientes. Sólo de este modo se podrán proponer métodos y soluciones que sean aplicables a cada uno de los casos. En este proyecto, se han utilizado las cohortes de UK Biobank y de GCAT. UK Biobank, es una cohorte de datos europeos que contiene la información genética, de estilo de vida y clínica de 500,000 individuos británicos. Por otro lado, GCAT, es una cohorte de datos española que contiene la información genética, de estilo de vida, ambiental y clínica de 20,000 participantes.

Uno de los grandes retos para poder analizar los datos de pacientes, es la sensibilidad de los datos con los que se trabaja. Por eso, es necesario establecer unos protocolos que faciliten la transferencia o acceso a los datos para poder ser analizados. En general, este aspecto sigue siendo uno de los máximos obstáculos en este tipo de investigaciones multidisciplinarias y transinstitucionales, y se propone como uno de los problemas aún no solucionados de la investigación médica. En este caso, se ha trabajado en colaboración con la UPV para supervisar y asesorar la generación de un Acuerdo de Transferencia de Datos (DTA) general entre BSC y IGTP, que pueda facilitar el acceso y la transferencia de datos entre los dos centros en este proyecto y en proyectos futuros. Este tipo de acuerdos, que suelen tener un formato estándar a nivel europeo, en el marco nacional todavía no están completamente estandarizados y, por tanto, el acceso de los centros de investigación básica a los datos de los pacientes sigue siendo un reto. Así, para poder acceder a la cohorte de UK Biobank, se realizó una solicitud de análisis que fue revisada por un comité, se firmó un DTA y previo pago, los datos fueron accesibles para el análisis. Y, por otro lado, en el marco de este proyecto y bajo la colaboración de IGTP, BSC y UPV, se ha conseguido no sólo el acceso a los datos de GCAT para probar los resultados obtenidos en el proyecto, sino también avanzar en la generación de este tipo de acuerdos y en la consolidación de protocolos que faciliten el acceso a los datos de los pacientes y del sistema público de salud. Otro de los grandes retos a los que se enfrentaba este proyecto es la estructuración de los datos. En general, todavía no se han definido unos estándares a nivel mundial que faciliten la comparación entre dos cohortes. De hecho, actualmente, el sistema clínico-sanitario no sigue unas reglas para automatizar el proceso de almacenamiento de los datos de los pacientes. En los últimos años, con el avance de las tecnologías y de las bases de datos y, conscientes de la necesidad de una armonización de los datos para el análisis conjunto, se han hecho grandes avances en esta dirección. De hecho, recientemente, han surgido distintas iniciativas para la creación de redes federadas de repositorios normalizados en Europa, con la intención de implantar el sistema OMOP. Sin embargo, estas iniciativas son tan recientes que todavía queda un largo camino para conseguir una

normalización de los datos. En este proyecto, en el BSC se han preparado los datos de UK Biobank para poder analizar los efectos de las variables genéticas, clínicas y de estilo de vida en el desarrollo de las enfermedades no comunicables, como la DT2, causante del 65.3% de las muertes a nivel mundial en 2021. Por otro lado, en IGTP se han conseguido identificar decenas de variables clínicas y de estilo de vida de GCAT con sus correspondientes de UK Biobank. A su vez, y en paralelo, este proyecto ha estado colaborando con los proyectos DATOSCAT e IMPaCTT2D para la armonización de los datos de la cohorte GCAT. Los avances en este sentido son fundamentales para el correcto análisis de los datos en este proyecto y para otros proyectos futuros.

En particular, la preparación de los datos de UK Biobank ha sido esencial para la creación de modelos de aprendizaje que faciliten la detección precoz de DT2. A su vez, la preparación de los datos de GCAT, y la identificación de las variables de estilo de vida y clínica de esta cohorte con las de UK Biobank, ha sido fundamental para evaluar la fiabilidad de los modelos predictores y para analizar la reproducibilidad de los resultados en cohortes independientes. Es decir, para asegurar que los modelos de DT2 que se han generado en este proyecto son aplicables para la detección precoz de la enfermedad en la población.

Para generar estos modelos predictores, hay que tener en cuenta que la DT2 es una enfermedad que se desarrolla por la combinación de múltiples variables genéticas que, desde nuestro nacimiento, predisponen a cada individuo a su desarrollo, en conjunto con la acción de distintos factores ambientales y de estilo de vida. Es decir, un efecto del entorno y de nuestros hábitos prolongados a lo largo del tiempo, junto con la predisposición genética de cada individuo es lo que genera la enfermedad. En el BSC, se han utilizado los datos de UK Biobank para obtener tres predictores distintos de DT2: genético, clínico y estilo de vida (ambiental). Cada uno de estos predictores es capaz de hacer una detección de la enfermedad en base a los datos correspondientes de cada paciente. Estos modelos predictores han resultado tener cada uno un poder de detección de la enfermedad de entre el 60 y el 73%. Con la ayuda de IGTP, se han probado estos modelos en la cohorte de GCAT y validado su robustez y fiabilidad.

A nivel genético, en comparación con los predictores que se utilizan actualmente, se ha observado que el uso de los modelos propuestos en este proyecto suponen un incremento en la detección de la predisposición genética al desarrollo de DT2. Además, tras una inspección de los marcadores genéticos seleccionados por el modelo, se observan posibles candidatos objetivos de estudio biológico. Por otro lado, las capacidades de predicción de los factores clínicos y ambientales se encuentran por debajo de los actuales modelos de diagnóstico de DT2. Esto es algo esperado ya que se pretende hacer una detección precoz de la enfermedad en base a su exposición a distintos factores de riesgo para corregir esos hábitos y retrasar e incluso evitar el desarrollo de la enfermedad. En esta dirección, se han estudiado las variables clínicas y de estilo de vida seleccionadas por el método predictor y se han generado listas que permiten hacer recomendaciones a cada paciente en función de la predicción obtenida por el modelo. Sin embargo, este proyecto no sólo tenía como objetivo la generación de estos modelos predictores sino la creación de un modelo predictor que combine todos estos factores. Con este fin, en el BSC se han estudiado las distintas posibilidades para generar este tipo de modelo predictor y sus protocolos de prevención asociados. Tras el análisis de los datos, se ha generado un protocolo que permite combinar las predicciones de cada modelo para computar un predictor de riesgo general de DT2. Por tanto, este predictor de DT2 contempla no sólo la predisposición genética del individuo a la enfermedad, sino también los efectos del estilo de vida del individuo y su reflejo a nivel clínico.

Además, este estudio ha revelado que ciertos patrones epigenéticos, específicamente en los sitios CpG, podrían estar vinculados al desarrollo de la diabetes tipo 2 (T2D). La mayoría de estos sitios se encontraron en regiones genómicas llamadas OpenSea, que, aunque no están directamente relacionadas con los promotores de los genes, podrían influir en su regulación. Entre los genes más destacados, se identificó TP53, asociado con la respuesta al estrés celular y la resistencia a la insulina, y TXNIP, clave en la regulación de la glucosa, que ya se sabía estaba sujeto a regulación epigenética. Estos hallazgos sugieren que factores externos, como la contaminación del aire, podrían estar promoviendo cambios epigenéticos mediante mecanismos como el estrés oxidativo, que afectan la homeostasis celular, afectando vías metabólicas relacionadas con la diabetes a través de estos genes y las redes genéticas implicadas. Este conocimiento abre nuevas posibilidades para predecir y prevenir la T2D, no solo enfocándose en tratamientos clínicos, sino también en la mitigación de los impactos ambientales en nuestra salud.