



整合平台示範計畫基因資料與 醫療大數據之整合

2022國家級人體生物資料庫整合平台第二屆年會報告

楊奕馨

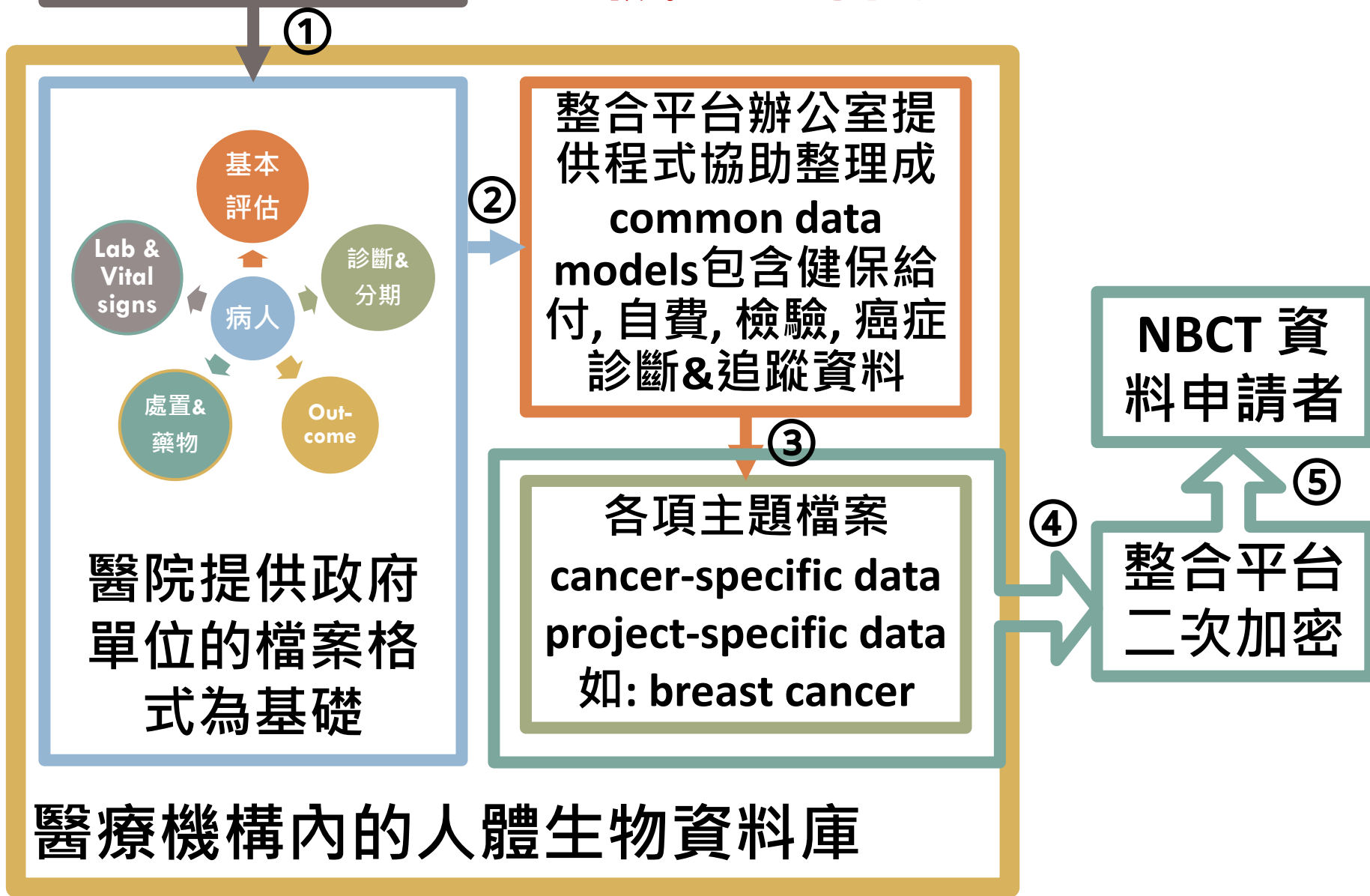
國家衛生研究院 癌症研究所

[Email: yhyang@nhri.edu.tw](mailto:yhyang@nhri.edu.tw)

<http://nicr.nhri.org.tw/pi/yangyh-cv/>

電子健康資料

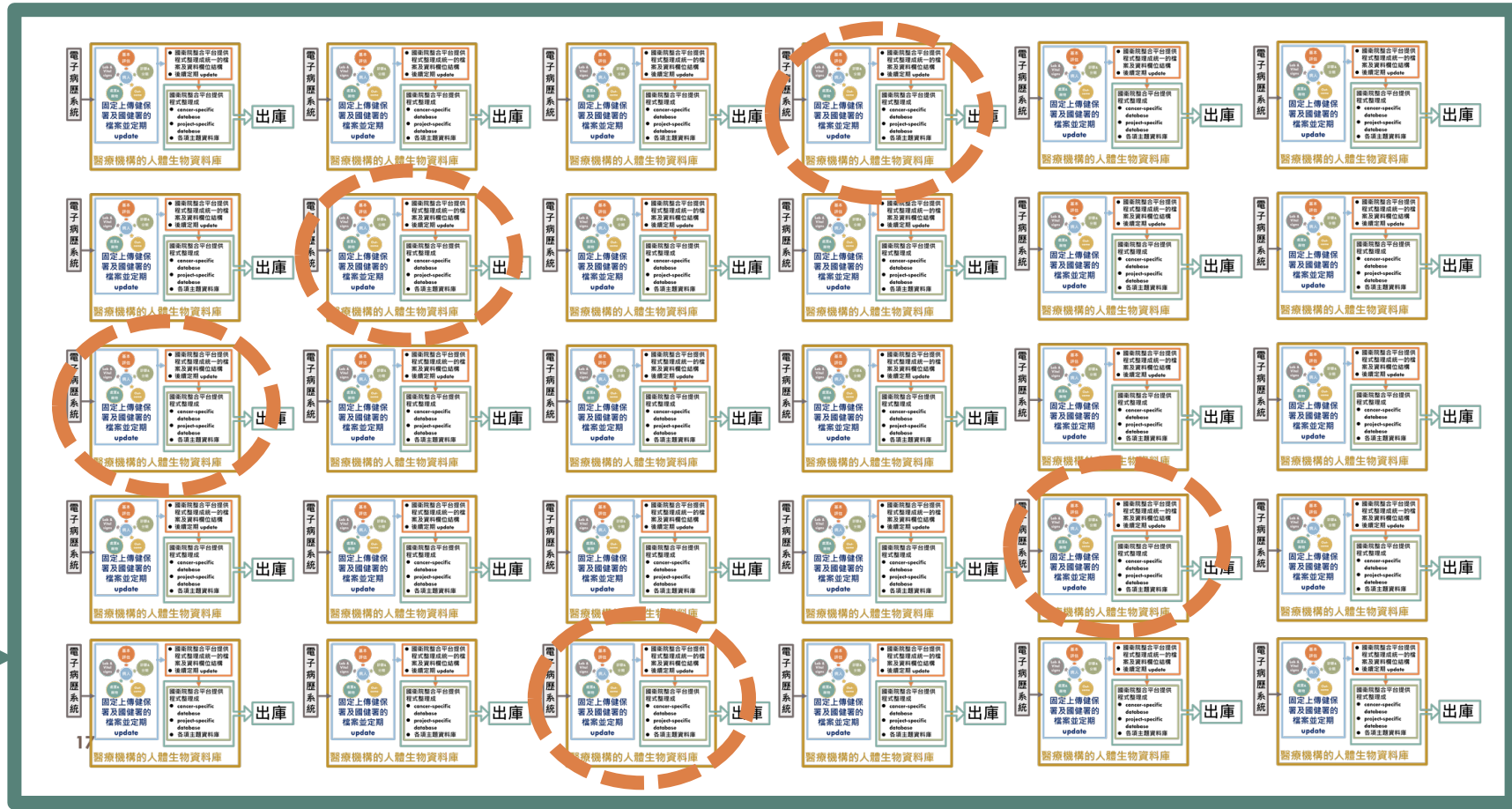
各機構內的NBCT CDM系統



整合平台與各參與Biobank之整合串聯情形

申請者向 NBCT 提出數據資料申請

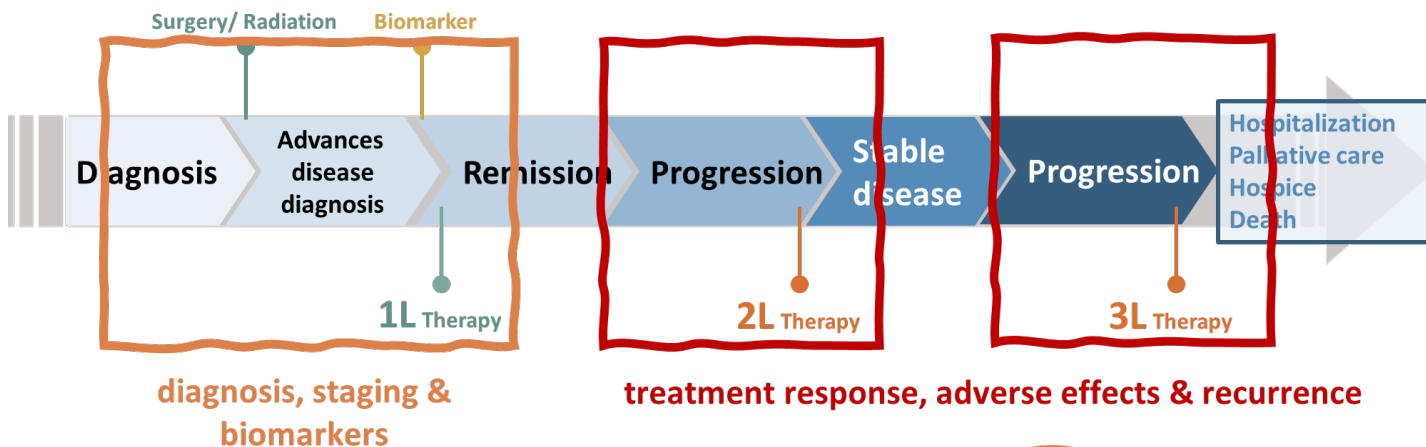
NBCT 聯繫醫院，進行醫院端流程



數據資料提供申請者使用

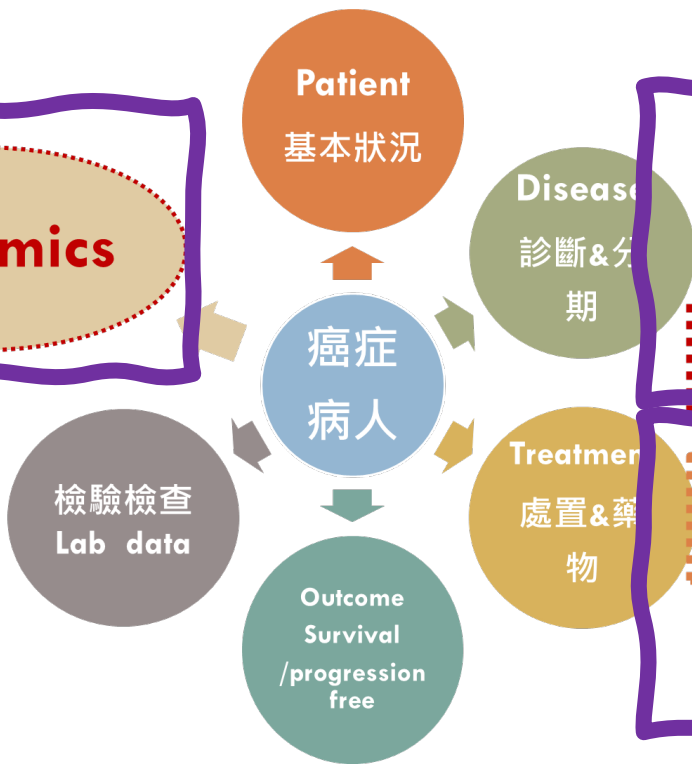
NBCT 彙整資料並二次重新編碼

示範計畫主題資料庫建置



收集病人治療歷程的資料足夠提供精準醫療的研究使用

FMI genomics reports
Genomics



NBCT CDM

臨床醫療資料 (EHR):

- 結構式資料 (第一階段)
- 非結構式資料 (後續)
 - 文字報告
 - 影像

尚未有結構欄位資料的資訊以人工病歷摘錄的方式登錄

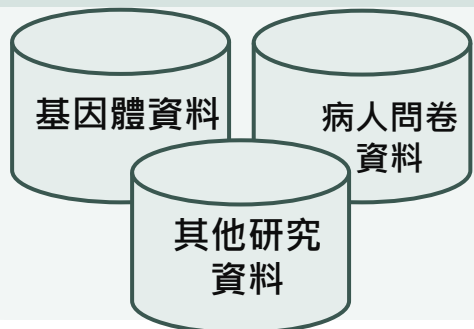
Linking EHR and NBCT data for precision medicine research

下午李修安博士會有系統設計的詳細說明

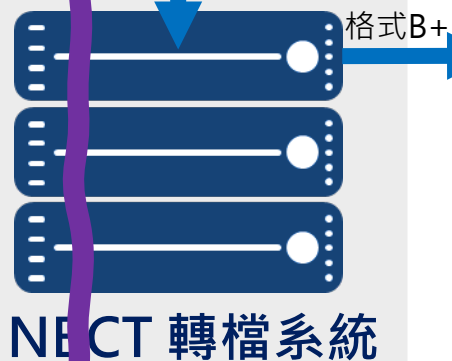
1. 臨床診療資料擷取



2. 外部資料篩選



3. Biobank 收案登錄資料篩選



- 主要癌症與疾病主題資料欄位擷取
- 外部資料匯入
- 資料篩選
- 基本描述統計視覺化

配合通過的案件出庫

6

三種來源資料的彙整

PATIENT
DISEASE Unknown primary melanoma
NAME
DATE OF BIRTH
SEX Female
MEDICAL RECORD #

PHYSICIAN
ORDERING PHYSICIAN
MEDICAL FACILITY
ADDITIONAL RECIPIENT
MEDICAL FACILITY ID
PATHOLOGIST

SPECIMEN
SPECIMEN SITE
SPECIMEN ID
SPECIMEN TYPE
DATE OF COLLECTION
SPECIMEN RECEIVED

Human readable

NO REPORTABLE ALTERATIONS WITH COMPANION DIAGNOSTIC (CDx) CLAIMS

See professional services section for additional information

OTHER ALTERATIONS & BIOMARKERS IDENTIFIED

Results reported in this section are not prescriptive or conclusive for labeled use of any specific professional services section for additional information.

```
<Sample>
<FM_Id>ORD-0782609-01</FM_Id>
<SampleId>US0775147.01</SampleId>
<BlockId>P00029</BlockId>
```

Machine readable (XML file)

	A	B	C	D	E	F	G
1	ReportId	Gene Name	Alteration Name				
2	QRF021632	KDR	amplification				
3	QRF021632	KIT	amplification				
4	QRF021632	PDGFRA	amplification				
5	QRF021632	RET	E511K				
6	QRF021632	BRAF	amplification				
7	QRF021632	CDKN2A/B	p14ARF loss exon 1 and CDKN2B loss				
8	QRF021632	Microsatellite status	MS-Stable				
9	QRF021632	TET2	Q1466*				
10	QRF021632	TP53	V172D				
11	QRF021632	Tumor Mutation Burden	TMB-Low				
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							

- Genomic findings
- Biomarkers finding
- Short variants
- Copy number alterations
- Microsatellite instability

GENOMIC FINDINGS

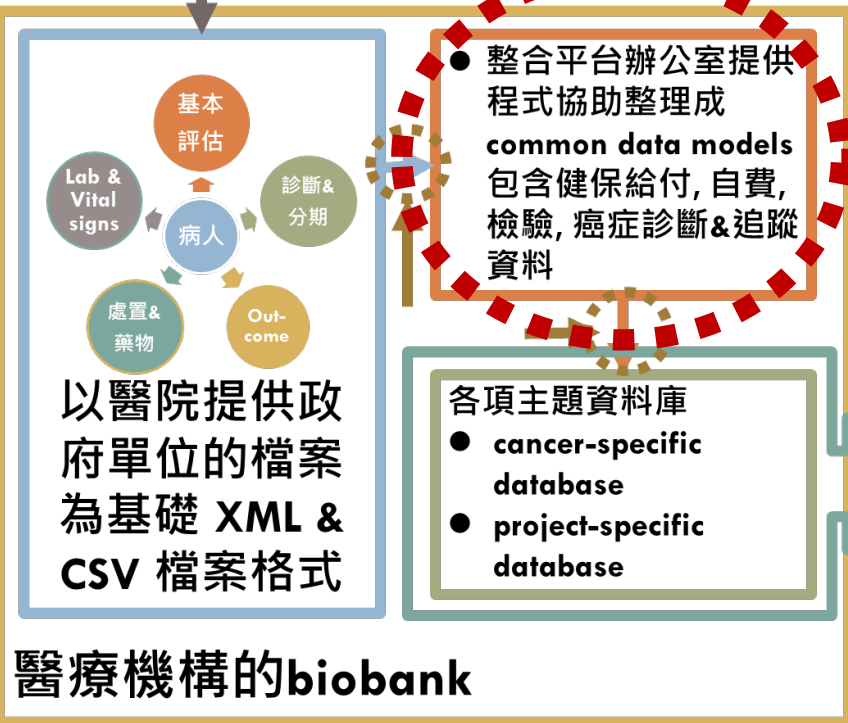
proliferation in a melanomas of 11.4 melanoma levels of VEGFR-2 with control related with

epithelial growth factor receptor tyrosine kinase VEGF-A and is development. KDR

in sun-damaged melanoma. One related KIT mRNA expression with KIT

on a cell surface leads to activation MAPK signaling, including point mutation, and/ associated with KIT being cancer and may

醫院電子數據資料



藍色上傳健保署資料, 紫色上傳國健署, 紅色自費項目

- 門診醫療申報檔案總表 (TOTFAE)
- 門診醫療申報檔案健保給付醫令 (TOTFAO1)
- 門診醫療申報檔案自費項目醫令 (TOTFAO2)
 - 自費項目以癌症藥品為主, 如附件:自費項目.pdf
- 住院醫療申報檔案總表 (TOTFBE)
- 住院醫療申報檔案健保給付醫令 (TOTFBO1)
- 住院醫療申報檔案自費項目醫令 (TOTFBO2)
 - 自費項目以癌症藥品為主, 如附件:自費項目.pdf
- 檢驗檢查資料檔格式(包含健保給付及自費項目) (LABH1 & LABH2)

- 癌症登記長表資料檔 (檔案代碼:CRLF)
- 癌症登記短表資料檔 (檔案代碼:CRSF)
- 死亡資料檔 (檔案代碼:DEATH)
- 癌症個案管理追蹤資料檔 (檔案代碼:CASE)

20

NBCT CDM 規劃足夠提供精準醫療資訊的臨床治療欄位, 並兼顧個資保護, 落實去識別避免再度辨識:

- ID's 進行hash處理 (資料分析時能夠內部串連)
- 日期模糊化處理

示範計畫 外部資料匯入

NBCT轉檔系統

首頁

資料匯入及轉檔

門診

住院

檢驗檢查

癌登表

死亡資料

個案追蹤

示範計畫

資料管理

示範計畫

示範計畫(001)檔案

FMI (genomic report) XML

整個資料夾上傳

Filename(檔案上限為200mb): 選擇檔案 未選擇任何檔案

送出

Access (臨床資料)

Filename(檔案上限為200mb): 選擇檔案 未選擇任何檔案

送出

ORD-0000000-01FL1CDx
XML 檔案
155 KB

示範計畫 - 病歷資料人工抄錄表單
Microsoft Access Database
2.25 MB

- ① 外部資料FMI report (XML file) 、 Case Report Form (ACCESS file)匯入NBCT轉檔系統，
- ② 系統擷取示範計畫收案病人的CDM後，
- ③ 產生完整的示範計畫癌症主題資料檔。

ID 關聯表

一鍵更新

最後更新資料(一天內)

總收錄資料

疾病主題檔檢索(自選)

疾病主題檔主題式檢索(預設)

疾病主題檔歷史紀錄

示範計畫(自選)

示範計畫

選擇檔案 未選擇任何檔案

匯入

清空欄位

送出資料

主題資料檔(參考)

肺癌

胃癌

食道癌

胰臟癌

肝外膽管癌

膽囊癌

- ① 外部資料 **FMI report (XML file)**、**Case Report Form (ACCESS file)** 匯入 NBCT轉檔系統，
- ② 系統擷取示範計畫收案病人的 **CDM** 後，
- ③ 接著匯出產生完整的 **示範計畫癌症主題資料檔**。

請輸入HashIL

請輸入ChartIL

chartnumber

示範計畫 機構資料出庫譯碼簿 (含CDM及case report form)

癌症精準醫療及生物資料庫整合平台

合作示範計畫

癌症主題資料檔 檔案格式及資料描述

(Non-squamous NSCLC, Esophageal cancer, Gastric cancer, Gallbladder cancer, Extrahepatic cholangiocarcinoma, Pancreatic cancer)

目錄

癌症主題資料檔 檔案格式及資料描述.....	4
1 Patient demographic 「from case report form (ACCESS file)」	4
1 Patient demographic 「from NBCT CDM」	4
2-1 Cancer Characteristic (Newly diagnosis) 「from case report form (ACCESS file)」	5
2-1 Cancer Characteristic (Newly diagnosis) 「from NBCT CDM」	5
2-2 Cancer Characteristic (First recurrence) 「from case report form (ACCESS file)」	7
2-2 Cancer Characteristic (First recurrence) 「from NBCT CDM」	8
3 Performance Status 「from case report form (ACCESS file)」	9
4 Comorbidities 「from NBCT CDM」	9
5 Laboratory results 「from NBCT CDM」	13
6 Biomarker test 「from NBCT CDM」	15
7-1 Cancer related medication (inpatient) 「from case report form (ACCESS file)」	16
7-1 Cancer related medication (inpatient) 「from NBCT CDM」	16
7-2 Cancer related medication (outpatient+ER) 「from case report form (ACCESS file)」	17
7-2 Cancer related medication (outpatient+ER) 「from NBCT CDM」	17
7-3 Cancer related medication (self-paid) 「from case report form (ACCESS file)」	18
7-4 Cancer related medication (free-of-charge) 「from case report form (ACCESS file)」	19
8-1 Comedication (inpatient) 「from NBCT CDM」	20
8-2 Comedication (outpatient+ER) 「from NBCT CDM」	21
9 Cancer related radiotherapy 「from case report form (ACCESS file)」	22
9 Cancer related radiotherapy 「from NBCT CDM」	22
10 Cancer related surgery 「from NBCT CDM」	23
11 Cancer related surgery 「from NBCT CDM」	24
12 Cancer treatment outcome 「from case report form (ACCESS file)」	25
13 Follow-ups 「from case report form (ACCESS file)」	25
14 Death data 「from NBCT CDM」	26
15 Value indicator of the project 「from case report form (ACCESS file)」	26
欄位譯碼.....	28
(一) 臨床期別組合-編碼定義表	28
(二) 臨床 T-編碼定義表	28
(三) 臨床 N-編碼定義表	29
(四) 臨床 M-編碼定義表	30
(五) 病理期別組合-編碼定義表	30
(六) 病理 T-編碼定義表	31
(七) 病理 N-編碼定義表	32
(八) 病理 M-編碼定義表	33
(九) 癌症遠端轉移部位-編碼定義表	33

癌症主題資料檔 檔案格式及資料描述

1 Patient demographic 「from case report form (ACCESS file)」

(一名個案僅一筆紀錄)

序號	欄位名稱	英文欄位名稱	屬性	長度	資料描述
1	病人 id (已去識別化)	hash_id	Char	32	個案已去識別化之 id，做為各檔案間串檔用
2	FMI 檢測編號 (已去識別化)	FMI_hash	Char	32	FMI 基因檢測報告上的編號做去識別化處理。一位個案對應一組 FMI 檢測編號，為串聯個案基因資料之鍵值。

1 Patient demographic 「from NBCT CDM」

(一名個案僅一筆紀錄)

序號	欄位名稱	英文欄位名稱	屬性	長度	資料描述
1	病人 id (已去識別化)	hash_id	Char	32	個案已去識別化之 id，做為各檔案間串檔用
2	生日年	Year_of_Birth	Char	4	YYYY：確認個案的出生年。 年不詳：9999。
3	性別	Gender	Char	1	1：男性 2：女性 3：其他 (如雙性人) 4：變性人 9：不詳或在病歷上未記載
4	身高(公分)	Height_cm	Char	3	記錄個案於首次治療前的身高。 000-099：< 100 cm 之身高數值 (採四捨五入) 100-998：≥ 100 cm 之身高數值 (採四捨五入) 999：病歷未記載或不詳。
5	體重(公斤)	Weight_kg	Char	3	記錄個案於首次治療前的體重。 000-099：< 100kg 之體重數值 (採四捨五入) 100-998：≥ 100kg 之體重數值 (採四捨五入) 999：病歷未記載或不詳。
6	醫院代碼(已)	Hosp_hash	Char	32	醫院之醫事機構代碼，已做去識別

示範計畫 機構資料出庫譯碼簿 (FMI report, 由.xml轉檔成.xlsx格式)

癌症基因檢測檔 欄位名稱及資料描述

(Non-squamous NSCLC, Esophageal cancer, Gastric cancer, Gallbladder cancer, Extrahepatic cholangiocarcinoma, Pancreatic cancer)

1 Report_Patient (一名個案僅一筆紀錄)

序號	英文欄位名稱	資料描述	欄位內容範例
1	FMI_hash	FMI's unique id for this sample (de-identified)	
2	SubmittedDiagnosis	Patient's diagnosis submitted with sample	Unknown primary melanoma
3	SpecSite	Location from which sample was retrieved	Rectum Blood
4	TestType	Name of the test performed on this sample	FoundationOneDx FoundationOneLiq uidDx

2 Report_genomic findings (一名個案多筆紀錄)

序號	英文欄位名稱	資料描述	欄位內容範例
5	FMI_hash	FMI's unique id for this sample (de-identified)	
6	Gene Name	Name of gene	KDR AR
7	Alteration name	Name of alteration	Amplification L702H

3 Report_biomarker findings (一名個案多筆紀錄)

序號	英文欄位名稱	資料描述	欄位內容範例
8	FMI_hash	FMI's unique id for this sample (de-identified)	
9	reportProperty	Property key	TumorMutationBurdenScore Microsatellite status
10	value	Value of defined property	3 Muts/Mb MS-Stable

20221121151833986267_F1CDx_GENE [唯讀] - Excel

檔案 常用 插入 頁面配置 公式 資料 校閱 檢視 說明 ACROBAT Power Pivot TreeAge Pro 告訴我您想做什么

剪下 剪貼簿 複製 複製格式 貼上

Calibri 11 A A 自動換行 通用格式 條件式格式設定 格式化為表格

B I U 對齊方式 數值

19

	A	B	C	D	E	F	G	H
1	ReportId	FMI_Id	SubmittedDiagnosis	SpecSite				
2	QRF021632	QRF021632	Unknown primary melanoma	Rectum				
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								

Report_Patient

20221121151833986267_F1CDx_GENE [唯讀] - Excel

檔案 常用 插入 頁面配置 公式 資料 校閱 檢視 說明 ACROBAT Power Pivot TreeAge Pro 告訴我您想做什么

剪下 剪貼簿 複製 複製格式 貼上

Calibri 11 A A 自動換行 通用格式 條件式格式設定 格式化為表格

B I U 對齊方式 數值

M10

	A	B	C	D	E	F	G	H	I	J
1	ReportId	Gene Name	Alteration Name							
2	QRF021632	KDR	amplification							
3	QRF021632	KIT	amplification							
4	QRF021632	PDGFRA	amplification							
5	QRF021632	RET	E511K							
6	QRF021632	BRAF	amplification							
7	QRF021632	CDKN2A/B	p14ARF loss exon 1 and CDKN2B loss							
8	QRF021632	Microsatellite status	MS-Stable							
9	QRF021632	TET2	Q1466*							
10	QRF021632	TP53	V172D							
11	QRF021632	Tumor Mutation Burden	TMB-Low							
12										
13										
14										
15										

Report_Patient Report_genomic findings Report_biomarker_findings Variant_short-variants Variant_copy-nun ...

後續

目前與各合作醫院進行NBCT CDM資料清理及出庫演練, 持續優化NBCT主題資料庫!



謝謝!

Questions & Comments?