

出國報告（出國類別：進修）

進行醫療研究，強化癌症治療品質與轉  
譯醫學應用的專業概念

服務機關：國立成功大學醫學院附設醫院外科部

姓名職稱：徐慧萍 副教授兼主治醫師

派赴國家：美國

出國期間：2019/8/1 至 2020/8/4

報告日期：2020/8/21

## 摘要

范德堡大學醫學中心是美國著名的醫學院與醫學中心，其中生物統計系肩負著教學、研究與服務的重任，是醫學中心裡舉足輕重的部門。此次在范德堡大學醫學中心生物統計系系主任石瑜教授的安排之下，本人有幸到范德堡大學醫學中心進修一年。除了在生物統計的部分有更深入的學習，也學習到研究相關資源的設立，更直接參與匿名病歷資料庫與基因資料庫的使用。藉由此次的進修，及與相關領域專家一同討論的經歷，不僅加強癌症治療與轉譯醫學的專業概念、提升研究經驗與擴展視野，並增加一般外科與乳房外科等癌症治療相關單位醫療品質與醫學研究的專業概念，也更將視野放置到國家級的資料庫與相關資源。

## 目次

壹、目的.....	第 1 頁
貳、過程.....	第 1-8 頁
參、心得.....	第 8-10 頁
肆、建議事項.....	第 10-11 頁
附錄.....	第 12 頁

## 壹、 目的：

根據國家健康署與衛生福利部的統計資料，惡性腫瘤現在已經成為臺灣人民死亡原因中的第一名，而乳癌是臺灣女性發生率最高的癌症，於民國 105 年在每十萬臺灣女性人口中有 15 018 人罹患乳癌，好發在 45 歲至 65 歲的女性同胞中；雖然現在有多元與全面的癌症治療方法，因乳癌而死亡還是佔臺灣女性十大癌症死亡的第三名，侵襲性乳癌病人的五年存活率為 88.4%，但有更多女性在診斷後的五至十年後才延遲性復發，代表著乳癌對於臺灣女性的健康有更深遠的影響。

因此在教育部、成功大學與成大醫院的支持下，我得以到美國田納西州（Tennessee）納許維爾（Nashville）的范德堡大學醫學中心（Vanderbilt University Medical Center），向生物統計系（Department of Biostatistics）的石瑜教授（Professor Yu Shyr）學習，並由其安排向相關領域專家一同探討癌症治療與轉譯醫學的專業概念，藉由此行加強研究經驗與擴展視野，並增加一般外科與乳房外科等癌症治療相關單位醫療品質與醫學研究的專業概念。

## 貳、 過程

石教授目前擔任生物統計系系主任，負責生物統計系的教學、研究與服務多項業務，尤其是在癌症相關的議題上，涉獵甚多。因此在我進修的過程中，經由石教授的安排，學習到如何以生物統計（biostatistics）的技術來針對乳癌治療與轉譯醫學。逐一報告如下：

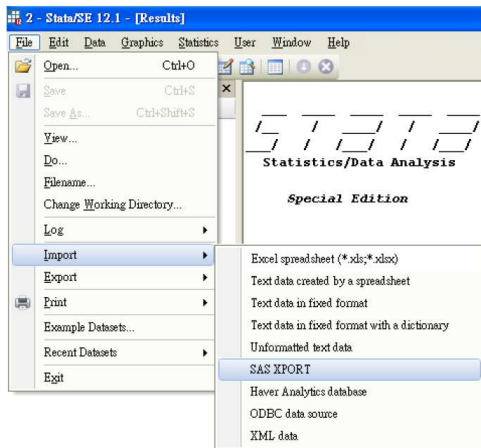
### 一. 教學：

癌症資料庫處理中，很重要的一部分是要藉由生物統計學來分析數據；更由於大數據（big data）及基因資料庫（gene database）的開發與應用，需要同時處理上百萬筆資料，生物統計與電腦科學（computer science）的重要性逐年增加。因此在我進修的第一部分，就是著重於關於生物統計的學習。

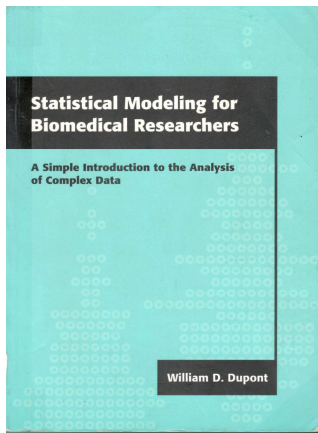
授課老師包括石瑜教授本人、Professor William D. Dupont、Dr. Qu Liu、Dr. Thomas Stewart 與 Dr. Fei Ye。授課內容包括生物統計簡介、製圖與製表、勝算比（Odds Ratio）與相對風險（Relative Risk）、卡方檢定（Chi-Squared Test）、

獨立樣本 t 檢定 (independent Student's t-test)、無母數分析 (non-parametric tests)、存活分析 (survival analysis)、簡單與多元線性迴歸分析 (simple and multiple linear regression)、簡單與多元邏輯迴歸分析 (simple and multiple logistic regression)、單變數分析 (univariate analysis)、及多變數分析 (multivariate analysis)。

經由這些課程的教授，我學習到 STATA 與 SPSS 軟體的應用，分析大數據及製備特殊圖形，並且能應用在臨床數據上，用於文獻發表，或者用作準備教學教材。適當的製圖與製表，有助於表現數據及提升學生學習效率。學習成效如下：



圖一、STATA 軟體之使用介面



圖三、Dupont 教授所撰之生物統計教科書

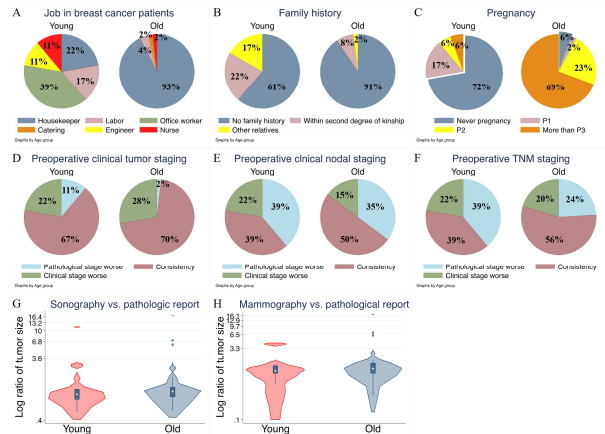
**Primary Instructors:**

**Yu Shyr, Ph.D.**  
 Harold L. Moses Chair in Cancer Research  
 Professor and Chairman, Department of Biostatistics  
 Director, Vanderbilt Center for Quantitative Sciences  
 Director, Vanderbilt Technologies for Advanced Genomics Analysis and Research Design  
 Professor of Biomedical Informatics, Health Policy, and Cancer Biology  
 Vanderbilt University Medical Center  
 576 Preston Research Building  
 2525 West End Ave, Ste. 1100  
 Work: (615) 936-2572/Cell: (615) 423-4943  
 Email: [yu.shyr@vumc.org](mailto:yu.shyr@vumc.org)

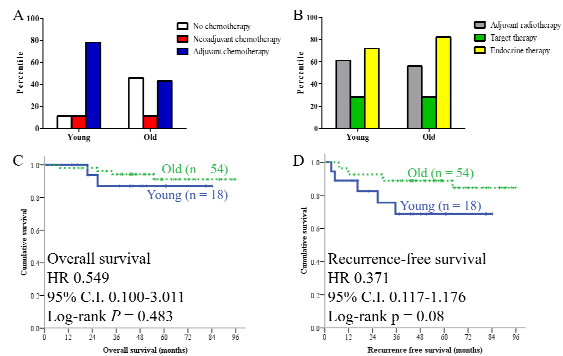
Assistant: Jill Shell  
 Work: (615) 936-6760  
 Email: [jillanne.k.shell@vumc.org](mailto:jillanne.k.shell@vumc.org)

Lecture	Date	Instructor	Topic	Due Dates
1	Thurs, 8/22	Shyr	Chapter 1: What is Biostatistics?	
2	Tues, 8/27	Shyr	Chapter 2: Introduction to Biostatistics	
3	Thurs, 8/29	Shyr	Chapter 3: Graphical Display of Data	HW 1
4	Tues, 9/3	Shyr	Chapter 4: Experimental Design, Sample Size Estimation and Power Analysis	
5	Thurs, 9/5	Shyr	Chapter 5: Probability Concept and Screening	HW 2
6	Tues, 9/10	Shyr	Review Lecture	

圖五、石瑜教授授課大綱



圖二、石瑜教授所授之臨床數據製圖



圖四、葉教授所授之存活分析製圖

Spring 2020

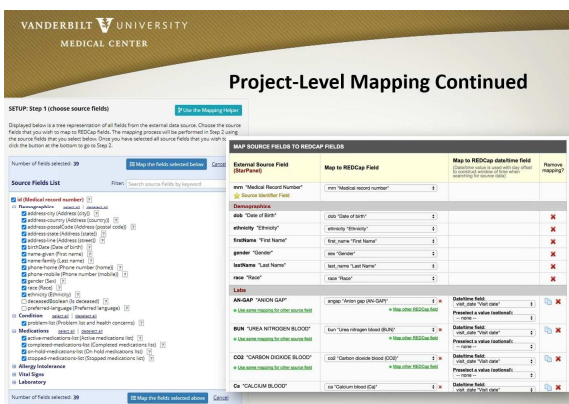
Month	Title	Presenter	After Seminar Student Discussion Leader
January 1	ICU SEVERITY - Winter break		
January 8	Data visualization (online) - Opportunities for the operationally-minded	Chris Johnson	Nil
January 15	Optimal Design of Two-Phase Studies	Alan Yu	Jeff Hsu, Pitt
January 22	Probability of Disease Changes in Academic Case-Finding Trials	Yueqin Zhou	Chang Yu
Feb. 5/week 24	Transmission-based infection in hospital: Infection control or negative ICU case?	Zach Hall	Nil
January 29	Polyploid cell counts: Ready for the clinic?	Jonathan Mosen	Nil
Feb/21			
February	Title	Presenter	After Seminar Student Discussion Leader
February 5	Integrated genetic methods for mutation studies in oncology research	Thomas Liorio	Jonathan Schmitt
February 12	Estimating Dose-to-Response in Case-Control Studies	Pai Lu	Ng Wing Sze/Edward
Feb. 19/week 14	Diagnostic Bayesian Network Model Using Regression Networks	Alan Au	Nil
February 19	10X single cell DNA sequencing data analysis from FACTO to differential expression analysis	Guangliang Sheng	Nil
February 26	Causal inference in prognostic randomized trials	Charles Murray	Sam Yee
Feb/28			

圖六、生物統計系演講

二. 研究：

乳癌相關研究是目前全世界各醫學中心與研發單位相當重視的一部分，我們成大醫學中心當然也非常重視乳癌研究。范德堡大學醫學中心是田納西州最大的醫學中心，其癌症研究成效卓著，有許多相關部門協助醫師與老師來進行研究工作，如：REDCap、VPRAD、VICTR、eMERGE 等。藉由此次進修的機會，我得以學習與了解到許多協助癌症研究的工具，適當使用這些工具有助於有效分析臨床數據及執行轉譯醫學。摘要重點如下：

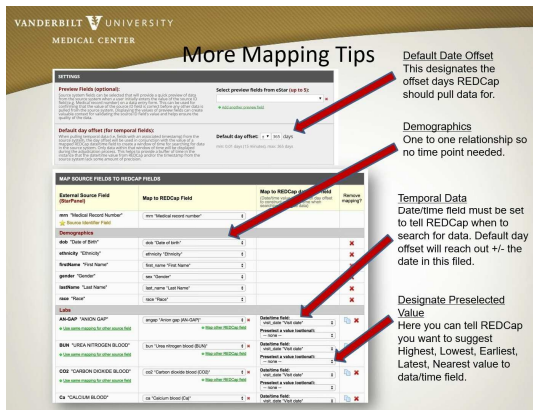
- ✧ REDCap：全名為 Research Electronic Data Capture，中文翻譯為臨床研究電子數據採集系統，是范德堡大學醫學中心 Professor Paul Harris 團隊所建立的系統，用於在大量電子病歷或是醫療數據庫中，針對研究者所需要的特定資訊，收集成統計相容軟體的資料庫。目前臺灣有多間醫學中心已經與范德堡大學醫學中心合作使用 REDCap 軟體，包括臺大醫院、長庚醫學中心、高雄醫學中心與我們成大醫院。但 REDCap 系統在判讀不同的電子病歷系統有不同的敏感度與專一性，對於中文病歷與英文病歷的解讀成功率差異更大，都還是目前需要克服的問題。



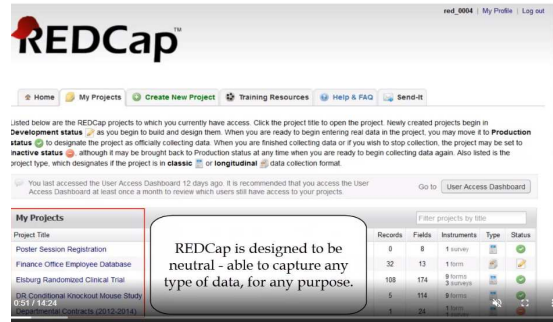
圖七、范德堡大學醫學中心 REDCap 使用介面 (一)



圖八、全世界各醫學中心與范德堡大學醫學中心合作使用 REDCap 之現況



圖九、范德堡大學醫學中心 REDCap 使用介面 (二)



圖十、REDCap 應用層次

- ✧ VPRAD：全名為 Vanderbilt University Medical Center Program in Research Administration Development，是范德堡大學醫學中心針對新進人員的職業發展計畫，協助尋找研究主題，或者針對已有的主題尋找范德堡大學醫學中心可以配合的資源或共同研究者。教學內容除了 12 周的正式課程（內含研究倫理、人體試驗、研究計畫等）之外，還包括有四個小時的工作坊，可以帶領研究者去了解動物中心、斑馬魚實驗室、基因定序中心、臨床研究中心等核心實驗室。



圖十一、VPRAD 是新進人員的職業發展



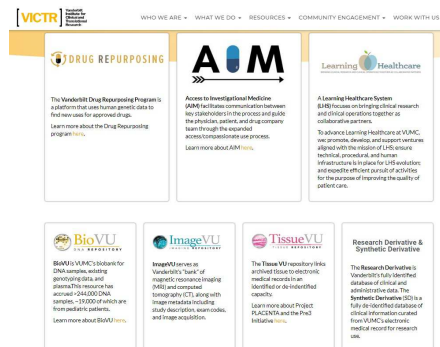
圖十二、VPRAD 工作坊

- ✧ VICTR：全名為 Vanderbilt Institute for Clinical and Translational Research，其主要工作是幫助醫師與老師將臨床或基礎研究的成果，應用在病人照護之上，是范德堡大學醫學中心內轉譯醫學的重要部門，有效幫助醫師與老師提升研究水準、增加發表文章、向外爭取經費、並且訓練年輕的研究者。VICTR 協同不同領域的研究者、提供各種訓練計

畫、提供經費補助、並且建立醫學資訊與生物統計系統，最後這部分包括很重要的匿名病歷資料庫（Synthetic Derivative）與基因資料庫（BioVU）。



圖十三、VICTR 業務範圍



圖十四、VICTR 相關規劃與子計畫

✧ eMERGE：全名為 Electronic Medical Records and Genomics (eMERGE) network，是由美國國家人類基因組研究所(National Human Genome Research Institute, NHGRI)與美國國家衛生研究院(National Institutes of Health, NIH)資助的研究團隊，其研究目標是為了要了解遺傳基因如何影響到人類健康，並且利用遺傳基因學研究成果來提供臨床上新醫療方法。目前已經找到多種疾病相關的遺傳基因型，或者是跟藥物反應有關的特殊遺傳基因，並且設立多個諮詢網站及知識庫，讓全世界的研究者都可以使用之。

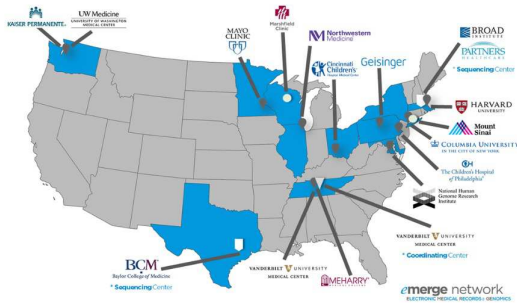
Disease Type	Genes Associated
Cancer	APC, ATM, BMPR1A, BRCA1, BRCA2, CHEK2, MEN1, MLH1, MSH2, MSH6, MUTYH, NF2, PALB2, PMS2, POLD1, POLE, PTEN, RB1, RET, SDHAF2, SDHB, SDHC, SDHD, SMAD4, STK11, TP53, TSC1, TSC2, VHL, WT1
Heart Disease	ACTA2, ACTC1, ANK2, APOB, CACNA1C, COL3A1, DSC2, DSG2, DSP, FBN1, GLA, KCNE1, KCNH2, KCNJ2, KCNQ1, LDLR, LMNA, MYBPC3, MYH7, MYH11, MYL2, MYL3, MYLK, PCSK9, PKP2, PRKAG2, RYR2, SCN5A, SMAD3, TGFBF1, TGFBF2, TMEM43, TNNT3, TNNT2, TPM1, TTR
Connective Tissue Disorders	COL3A1, COL5A1, FBN1, SMAD3, TGFBF1, TGFBF2
Metabolic Conditions	ACADM, ALDOB, ATP7B, BCKDHB, CPT2, CYP21A2, FAH, G6PC, OTC
Blood Disorders	F5, FANCC, HFE, JAK2
Other Conditions	BMPR2, BLM, CACNA1A, CACNA1S, HNF1A, HNF1B, MEFV, RYR1

圖十五、eMERGE 中與疾病相關的遺傳基因

Gene	Common Drugs Affected	More Information
CYP2C19	Clopidogrel	<a href="https://www.mydruggenome.org/dgi/clopidogrel/">https://www.mydruggenome.org/dgi/clopidogrel/</a>
	Voriconazole	<a href="https://www.mydruggenome.org/dgi/cyp2c19/">https://www.mydruggenome.org/dgi/cyp2c19/</a>
	Citalopram	
SLCO1B1	Escitalopram	
	Amitriptyline	
TPMT	Simvastatin	<a href="https://www.mydruggenome.org/dgi/simvastatin/">https://www.mydruggenome.org/dgi/simvastatin/</a>
CYP2C9	Azathioprine	<a href="https://www.mydruggenome.org/dgi/thiopurine/">https://www.mydruggenome.org/dgi/thiopurine/</a>
	Mercaptopurine	
VKORC1	Thioguanine	
	Warfarin	<a href="https://www.mydruggenome.org/dgi/warfarin/">https://www.mydruggenome.org/dgi/warfarin/</a>
	Warfarin	<a href="https://www.mydruggenome.org/dgi/warfarin/">https://www.mydruggenome.org/dgi/warfarin/</a>

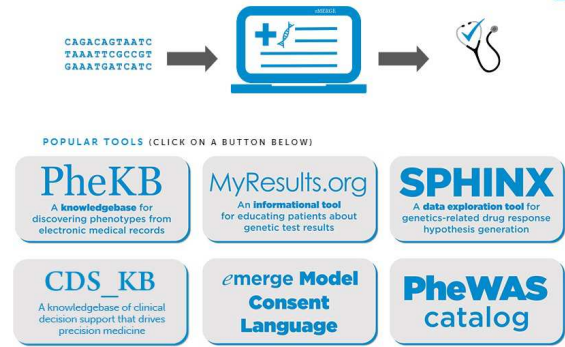
圖十六、eMERGE 中與藥物反應有關的特殊遺傳基因





圖十七、eMERGE 網路所包括的美國各個醫學中心

eMERGE is a national network organized and funded by the National Human Genome Research Institute (NHGRI) that combines DNA biorepositories with electronic medical record (EMR) systems for large scale, high-throughput genetic research in support of implementing genomic medicine.



圖十八、eMERGE 已設立的諮詢網站及知識庫

### 三. 服務：

除了基本的醫療服務、研究諮詢與教學任務之外，范德堡大學醫學中心最重要的兩項服務為匿名病歷資料庫（Synthetic Derivative）與基因資料庫（BioVU）。過去已經在石瑜教授於成大醫學院的演講，聽過這兩個資料庫的簡單介紹，但對於其應用與優缺點，了解多；但藉由此次進修的機會，我得以直接接觸到這兩項資料庫，適當使用這些資料庫有助於有效分析臨床數據及執行醫學研究，這其實是針對醫師與老師很重要的服務項目。摘要重點如下：

- ◇ SD：全名為 Synthetic Derivative，是范德堡大學醫學中心的匿名病歷資料庫。其最重要的部分，就是「匿名」。經由特殊運算方法（Secure Hash Algorithm），將電子病歷中的 18 項可能會推定出病人個人資料的項目移除、將時間用不規則的方式平移後，保留最多的資訊，以供內部的研究者使用。研究者必須接受人體研究與醫學倫理訓練，提出有具體目標的研究計畫，經由范德堡大學醫學中心的人體研究委員會審查，通過之後才能使用匿名病歷資料庫。

## Synthetic Derivative (SD) De-identification Process & Policy

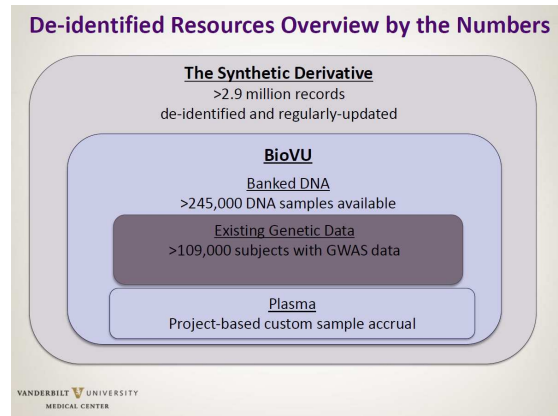
### List of 18 Identifiers

- Names
- All geographical subdivisions smaller than a state
- All elements of dates (except year) for dates directly related to an individual
- Phone numbers
- Fax numbers
- Electronic mail addresses
- Social Security numbers
- Medical record numbers
- Health plan beneficiary numbers
- Account numbers
- Certificate/license numbers
- Vehicle identifiers and serial numbers
- Device identifiers and serial numbers
- Web Universal Resource Locators (URLs)
- Internet Protocol (IP) address numbers
- Biometric identifiers including finger and voice prints
- Full face photographic images and any comparable images
- Any other unique identifying number, characteristic, or code

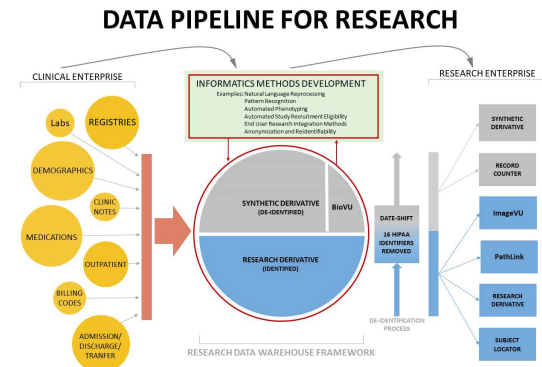
圖十九、SD 移除的個人資料主要項目



圖二十一、SD 介面

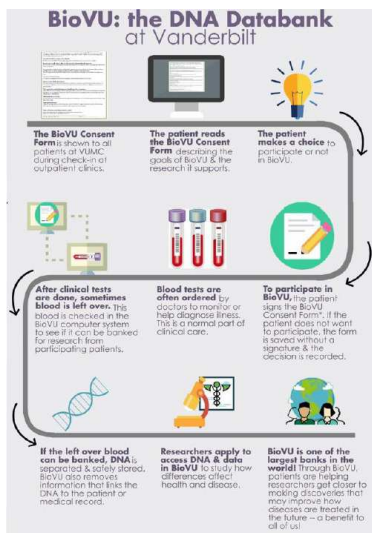


圖二十、圖解 SD 與 BioVU 包括的範圍

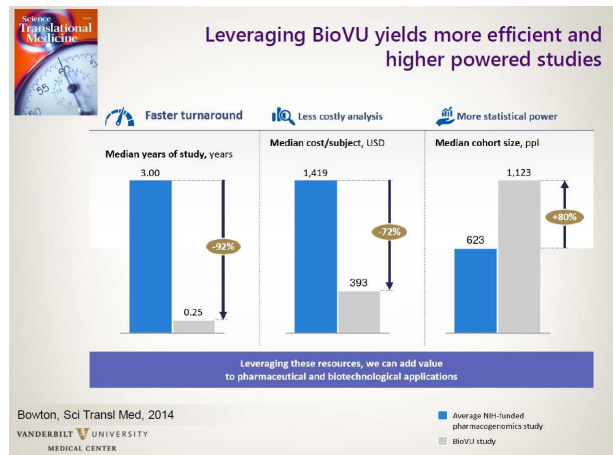


圖二十二、研究相關資料庫

◇ BioVU：是范德堡大學醫學中心的匿名基因資料庫，其最重要的部分，當然還是「匿名」。BioVU 是 SD 的延伸：所有范德堡大學醫學中心的病歷資料都是匿名後放入 SD 中，其中同意捐贈檢體的病人，其去氧核糖核酸 (deoxyribonucleic acid, DNA) 則收集入 BioVU 資料庫，而能與匿名的 SD 病歷資料庫串聯。若在 SD 資料庫中發現特殊疾病，可以利用 BioVU 資料庫分析可能相關突變基因；反之，若在 BioVU 資料庫分析特定的突變基因，可以在 SD 資料庫中找到病人可能會表現的特徵。因此 SD 與 BioVU 的相連結，提供給范德堡大學醫學中心研究者相當大的研究資源。執行 BioVU 基因定序的研究中心為 VANTAGE (Vanderbilt Technology for Advanced Genomics)，此次有幸去參觀 VANTAGE，見識到一流研究中心的研究設備與相關資源，的確是現代醫學研究很重要的一部分。



圖二十三、BioVU 匿名基因資料庫的檢體收集流程



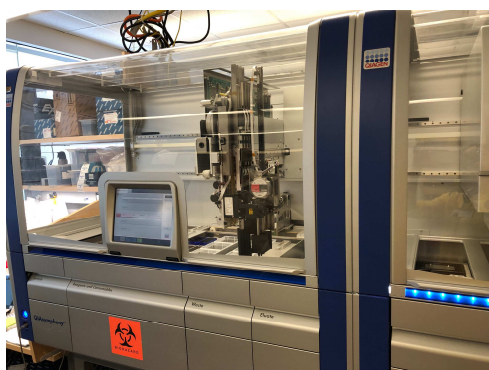
圖二十四、在 BioVU 系統引進後對於研究的具體幫助



圖二十五、VANTAGE 生物資料庫與電腦中心



圖二十六、基因定序儀 NovaSeq 6000



圖二十七、DNA 抽取操作平台

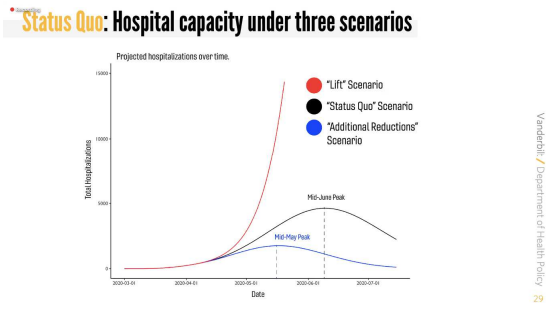


圖二十八、基因定序儀 NextSeq 500

## 參、心得：

此次到范德堡大學醫學中心進修一年，印象最深刻的就是 SD 匿名病歷資料庫與 BioVU 匿名基因資料庫對於研究所帶來的具體幫助。以 2019 年底開始全世界流行的新型冠狀病毒肺炎（COVID-19）為例，到 2020/4/8 為止范德堡大學醫學中心有 12 475 位病人接受反轉錄聚合酶鏈鎖反應檢測病毒核酸，陽性結果有





圖三十五、預測模型（三）

## 肆、建議事項

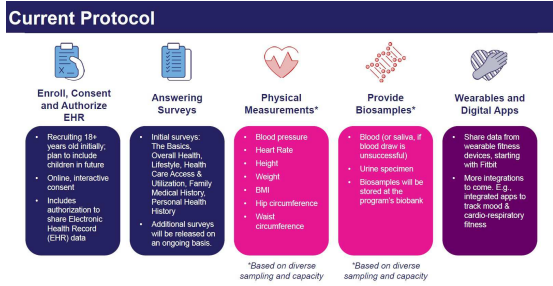
除了范德堡大學醫學中心自行開發的 SD 匿名病歷資料庫與 BioVU 匿名基因資料庫之外，美國國家衛生研究院也成立相關資料庫系統「All of US」，到 2019 年第四季已經有 224 350 位參與者的基本資料、其中 217 006 位的生活習慣紀錄、39 253 位的個人醫療紀錄、45 732 位的醫院醫療紀錄、與 108 640 位的電子醫療病歷。這樣資料庫的設立並不只是可以應用在乳癌研究，也可以應用在其他的疾病或公共衛生議題；並且，憑單一醫學中心的病人數與經費，會限制能夠設立資料庫的規模與深度；倘若能以國家級的設備與經費，設立全國性的資料庫，才能盡善盡美。臺灣目前的健保資料庫其實是全世界醫療資訊最完整、普及度最高的資料庫，但其去連結的方法是利用 k 匿名化 (k-anonymity)，在現實世界中是可能被破解的。因此建議政府落實匿名病歷資料庫的設計，並且涵蓋匿名基因資料庫，有助於協助解決醫療健康問題。



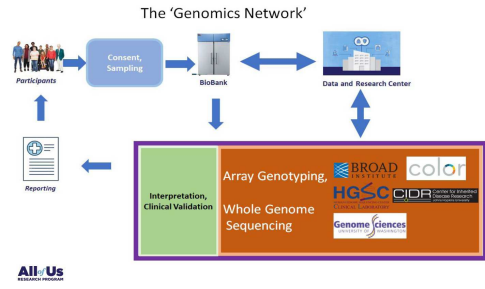
圖三十六、All of US 設立宗旨



圖三十七、評估方式



圖三十八、All of US 內容



圖三十九、基因收集與分析

附錄：

1. 中華民國 105 年癌症登記報告，衛生福利部國民健康署編印，中華民國 107 年 12 月出版。
2. William D. Dupont. Statistical Modeling for Biomedical Researchers: A Simple Introduction To The Analysis Of Complex Data. Cambridge Medicine.
3. [www.project-redcap.org](http://www.project-redcap.org)
4. <https://www.mc.vanderbilt.edu/myvumc/index.html?article=22840>
5. <https://victr.vumc.org/>
6. <https://medsites.mc.vanderbilt.edu/personalizedmedicine/emerge>
7. <http://vantage.vanderbilt.edu/>
8. <https://allofus.nih.gov/about>
9. <https://www.vumc.org/dbmi/synthetic-derivative>
10. <https://www.vumc.org/dbmi/biovu>