

# Enriching Health Science Alliance Biobank with clinical data from New South Wales Cancer Registry



Shum M<sup>1</sup>, Porwal M<sup>2</sup>, Barat S<sup>1</sup>, Granovskaya P<sup>1</sup>, Xu F<sup>2</sup>, Macmillan S<sup>2</sup>, Goldstein D<sup>2</sup>, Liauw W<sup>1</sup>

<sup>1</sup>South Eastern Sydney Local Health District, <sup>2</sup>Translational Cancer Research Network, Health Science Alliance Biobank, Prince of Wales Clinical School, UNSW Sydney



UNSW SYDNEY



Health South Eastern Sydney Local Health District

Cancer Institute NSW  
Translational Cancer Research Centre

## BACKGROUND

Biobanking plays a critical role in facilitating translational research. Biospecimens annotated with clinical data are used by researchers to examine causes, prevention, diagnosis, treatment and genetics of disease.

The New South Wales Cancer Registry (NSWCR) is the first Australian population-based cancer registry that captures data relating to episodes of care for people treated for cancer in NSW. Data from the Health Science Alliance (HSA) Biobank consented participants were submitted for clinical data linkage to NSWCR.

## AIMS

- To annotate biospecimens with clinical data from NSWCR for all participants consented for the year 2014.
- To map and link clinical data (diagnosis, grade, stage, treatment) available in NSWCR from the Case Repository level or Return to Notifier (RTN) notifications. For the purpose of the analysis, only one hospital site, Prince of Wales Hospital was considered.

## METHODS

- Data from all patients consented in 2014 & 2015 from the HSA Biobank were mapped and linked to NSWCR using key identifiers and data dictionaries
- Data fields were colour coded using a traffic light scheme

## RESULTS

Fig. 1: Linkage results NSWCR and HSA Biobank (2014)

### 2014

- ✓ Data linked in 149/201 (74%) cases
- ✓ Positive and robust data capture processes

Linked cases N	Description
31	Data found in NSWCR and the HSA Biobank (e.g. patient demographics, date of birth)
43	Partial data could be linked to the HSA Biobank due to data restricted at case level of NSWCR
75	Data restricted due to governance issues
52	Out of scope (i.e. diagnosed in 2005 or non-malignant cases)

- Full data access (21%)
- Partial data access (29%)
- Restricted data access (50%)

Fig. 2: Learnings from NSWCR & HSA data linkage (2014)

Gaps identified	Example	Potential solutions
Differences in data base design	Filed name- Gender or Sex	✓ Harmonise data fields prior to linkage
Data governance issues	Restricted data access at Return to Notification (RTN) and Case Repository level	✗ Apply for appropriate level of access ✓ Assess the data linkage needs
Eligibility criteria	Malignant vs. Non-malignant cases	✓ Filter patients based on eligibility criteria prior to

Legend ✓ Completed ✗ Under review

Learnings from 2014 data linkage are summarised in (Fig. 2). The HSA Biobank prioritised and re-defined the scope of the data fields leading to enhanced linkage (88%) of cases in 2015 between NSWCR & the HSA Biobank (Fig. 3).

### 2015

- ✓ Data linked in 79/90 (88%) cases
- ✓ Enhanced data capture processes

Fig. 3: Detailed linkage results NSWCR and HSA Biobank (2015)

Available data variables	Linked cases, N	Data access
Date of primary diagnosis,	60	Full
Tumour morphology <sup>^</sup> ,	30	Blank
Primary site		
Laterality <sup>^^</sup>	19	Full
	30	Blank
	41	Not applicable
Histopathological grade <sup>^^^</sup>	51	Partial
	30	Blank
	09	Missing

<sup>^</sup> Morphology is the histological classification and description of the course of development that a tumour can take as represented by a code

<sup>^^</sup>A paired organ is one in which there are two separate organs of the same kind, one on either side of the body

<sup>^^^</sup>The histopathological grade, differentiation or phenotype describes how little the tumour resembles the normal tissue from which it arose

## DISCUSSION

HSA Biobank was successfully enriched with clinical data via NSWCR data linkage project. Learnings from 2014 enabled enhanced and robust data access capabilities in 2015 linkage results. Majority of HSA Biobank cases were linked to the NSWCR. Some of the identified gaps i.e. lack of common identifiers, differences in database design were overcome for enhanced 2015 linkage. Preliminary learnings from this pilot project will be embedded into prospective data linkage.

## ACKNOWLEDGMENTS

The Health-Science Alliance  
SYDNEY • AUSTRALIA