Biases and ethics

* Genomic and pathway databases ingest the scientific literature

* The relative volume of literature in various fields is driven by funding

  * Databases are biased to cancer, heart disease and diabetes

  * Good luck if you have a rarer disease to treat

Biases and ethics

* Historically, minorities had lower chances of getting loans, mortgages, etc.

* This is reflected in the historical data

* This data is used to train ML models

* Now the historical racial bias is codified, since the models will make predictions based on patterns it saw in the training data

* Similar issue with clinical trials

  * Most studies were done on middle aged white men

  * Do these data apply to women, children, other ethnicities?

* Huge issue recognized in data ethics