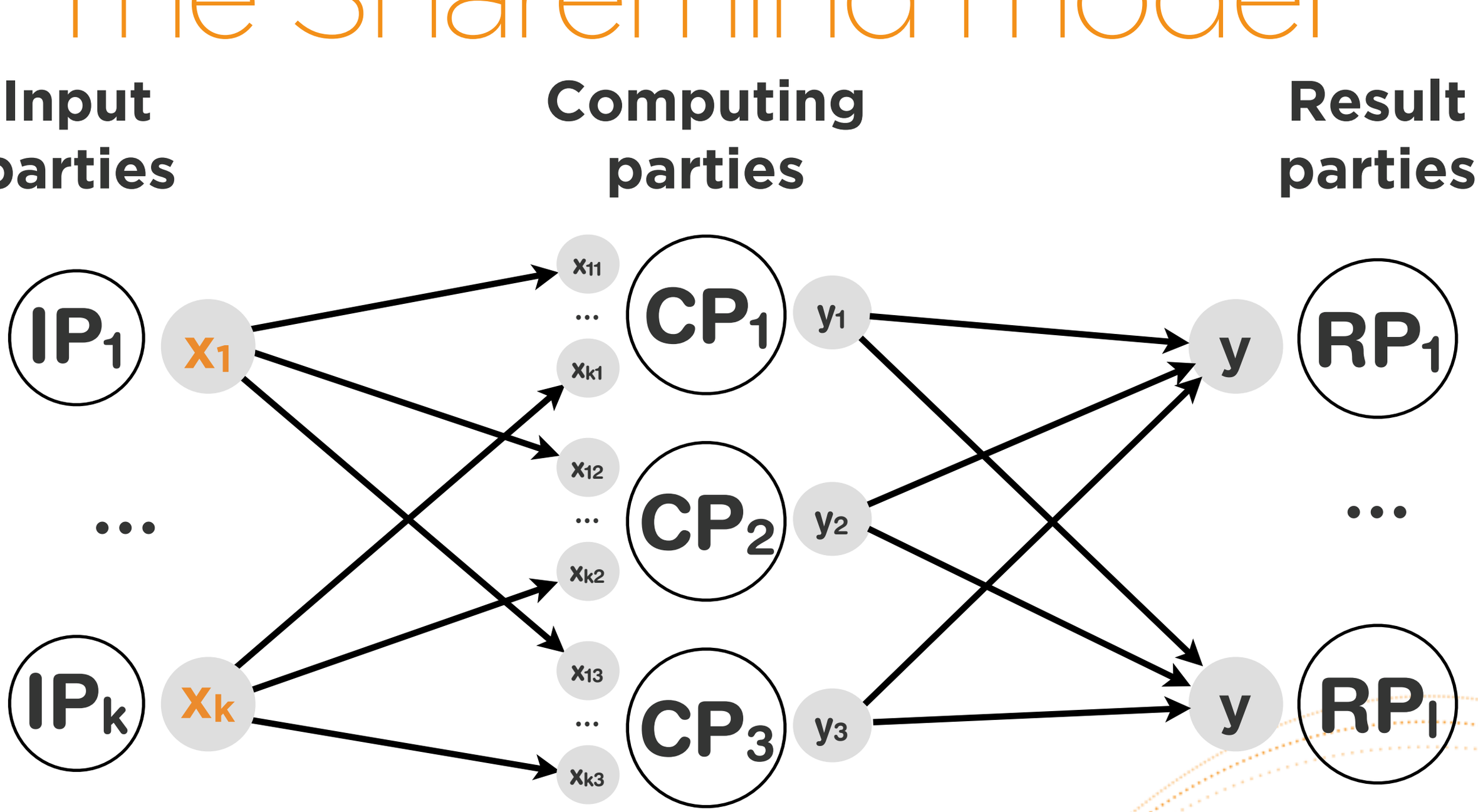
Full statistical analyses with secure multi-party computation

Dan Bogdanov, Liina Kamm, Ville Sokk

dan@cyber.ee
http://sharemind.cyber.ee/

sharemind

The Sharemind model



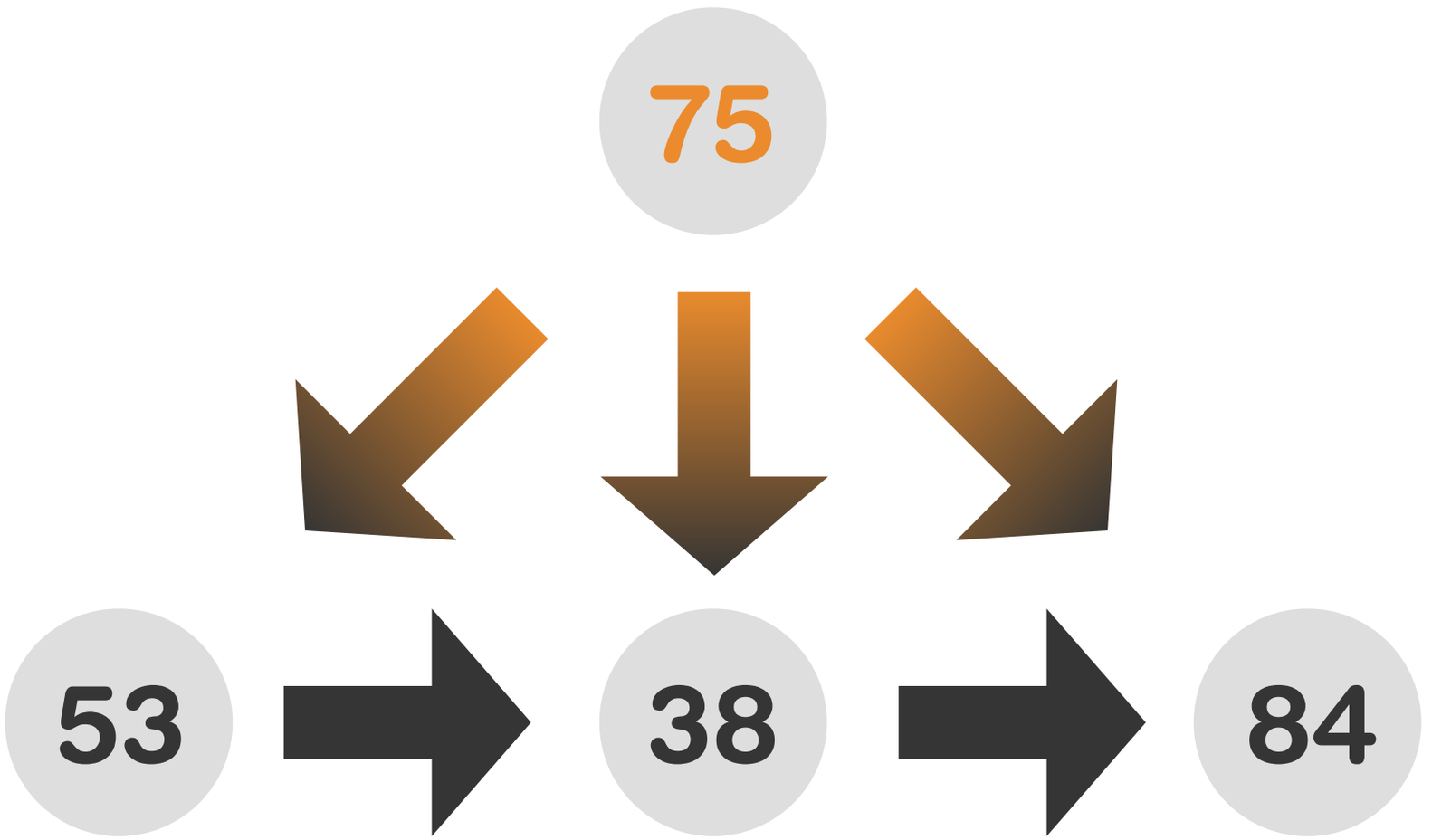
Step 1:
secret sharing and storage of inputs

Step 2:
secure multi-party computation

Step 3:
reconstruction of results

sharemind

Secret sharing (simplified)

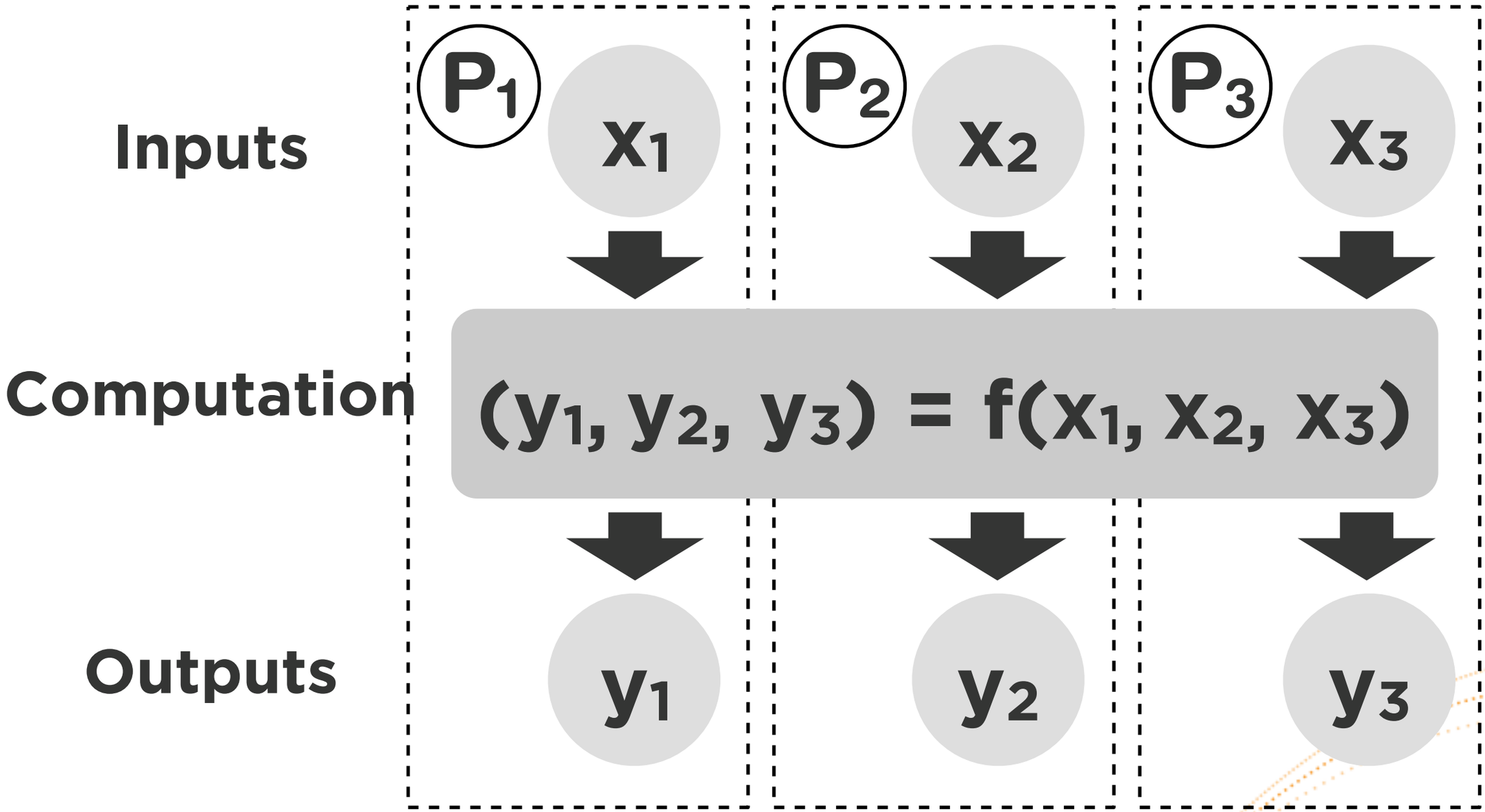


75 - 53 - 38 = 84 mod 100

Reconstruction: 53 + 38 + 84 = 75 mod 100

sharemind

MPC from secret sharing



All operations are composable.

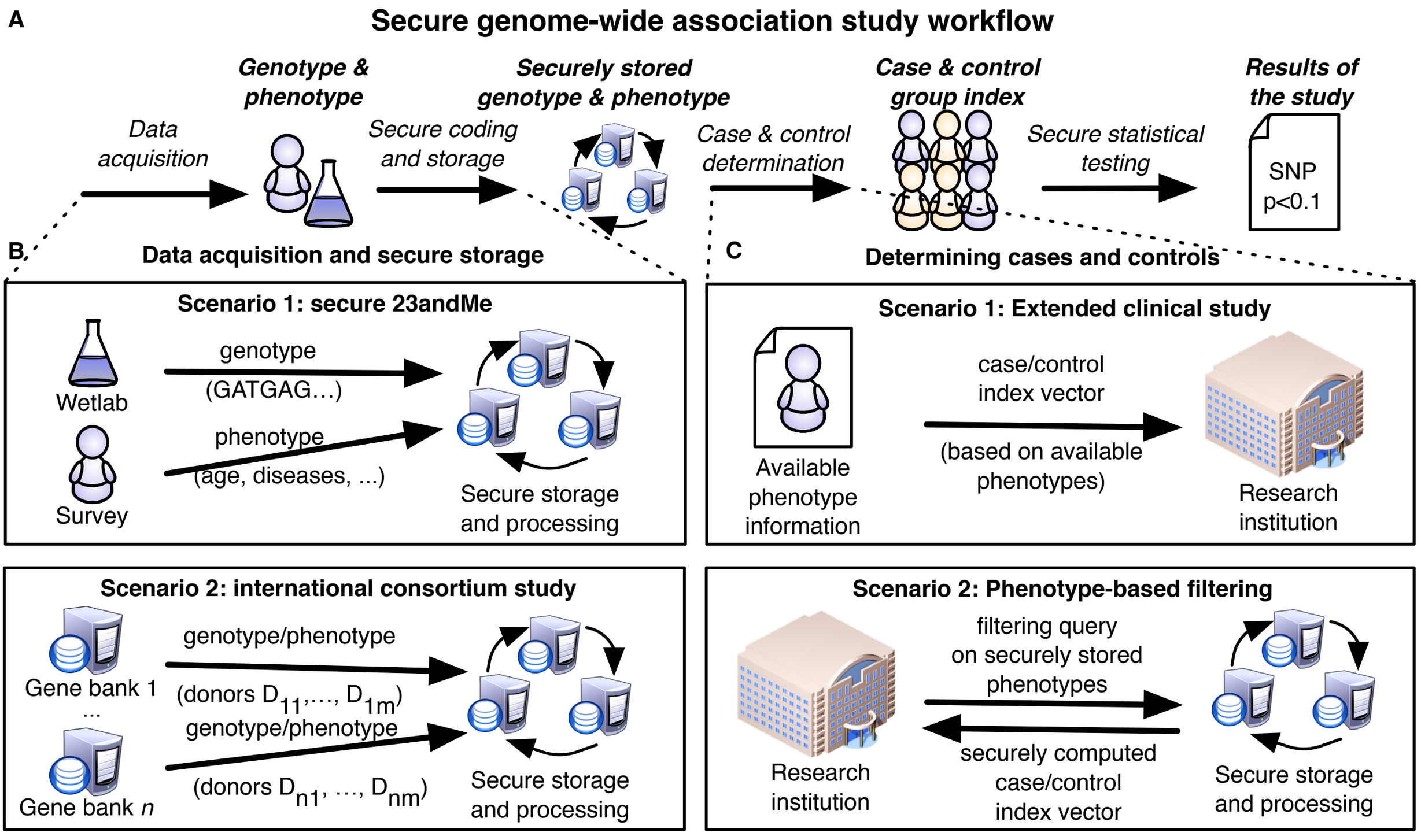
sharemind

Strengths / weaknesses

- **Easy to write code for.** Developers apply privacy patterns on classical algorithms.
- **Hybrid execution model** for balancing public and private computations.
- **Very high performance** for arithmetic circuits.
- **Small storage overhead** (3 times for 3 servers).
- **Requires three servers for best possible efficiency** (works with 2 to n servers as well).
- **Performance profile not immediately intuitive.**
- **Custom protocols** may perform better in some cases.

sharemind

Genome data and MPC



Application development

Description of the data analysis task

Business logic

SecreC language

Data model

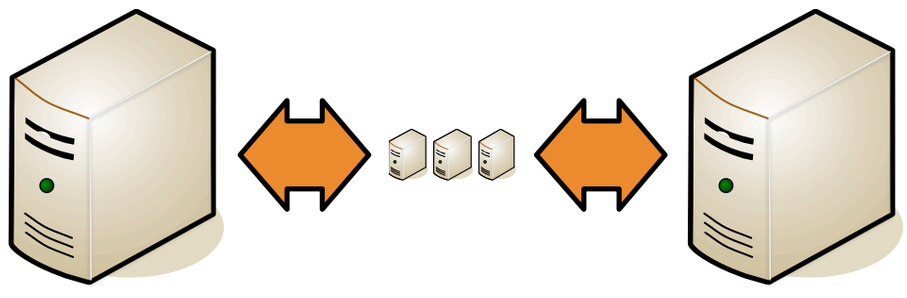
UX requirements

Controller library

Application Server package

sharemind

secure application servers



End user applications

end users (data owners, analysts etc)

Our competition entry

- **Task 2.1**
  - Importer (C++/SecreC), ~200 lines of code
  - Analyzer (C++/SecreC), ~200 lines of code
  - Secure operations used: secure integer arithmetic, floating point arithmetic, including division.
- **Task 2.2**
  - Importer (C++/SecreC), ~200 lines of code
  - Analyzer (C++/SecreC), ~300 lines of code
  - Secure operations used: secure integer arithmetic, shuffling, AES.

sharemind

The Rmind tool

LXTerminal

File Edit Tabs Help

LXTerminal | LXTerminal

```
'citation()' on how to cite R or R packag

Type 'demo()' for some demos, 'help()' fo
'help.start()' for an HTML browser interf
Type 'q()' to quit R.

> subject <- read.csv ("subject1000.csv",
> salary <- read.csv ("avg-salaries.csv",
> edu <- merge (subject, salary)
> age <- edu$age
> sal <- edu$avgSalary100
> plot(age, sal)
>
```

LXTerminal

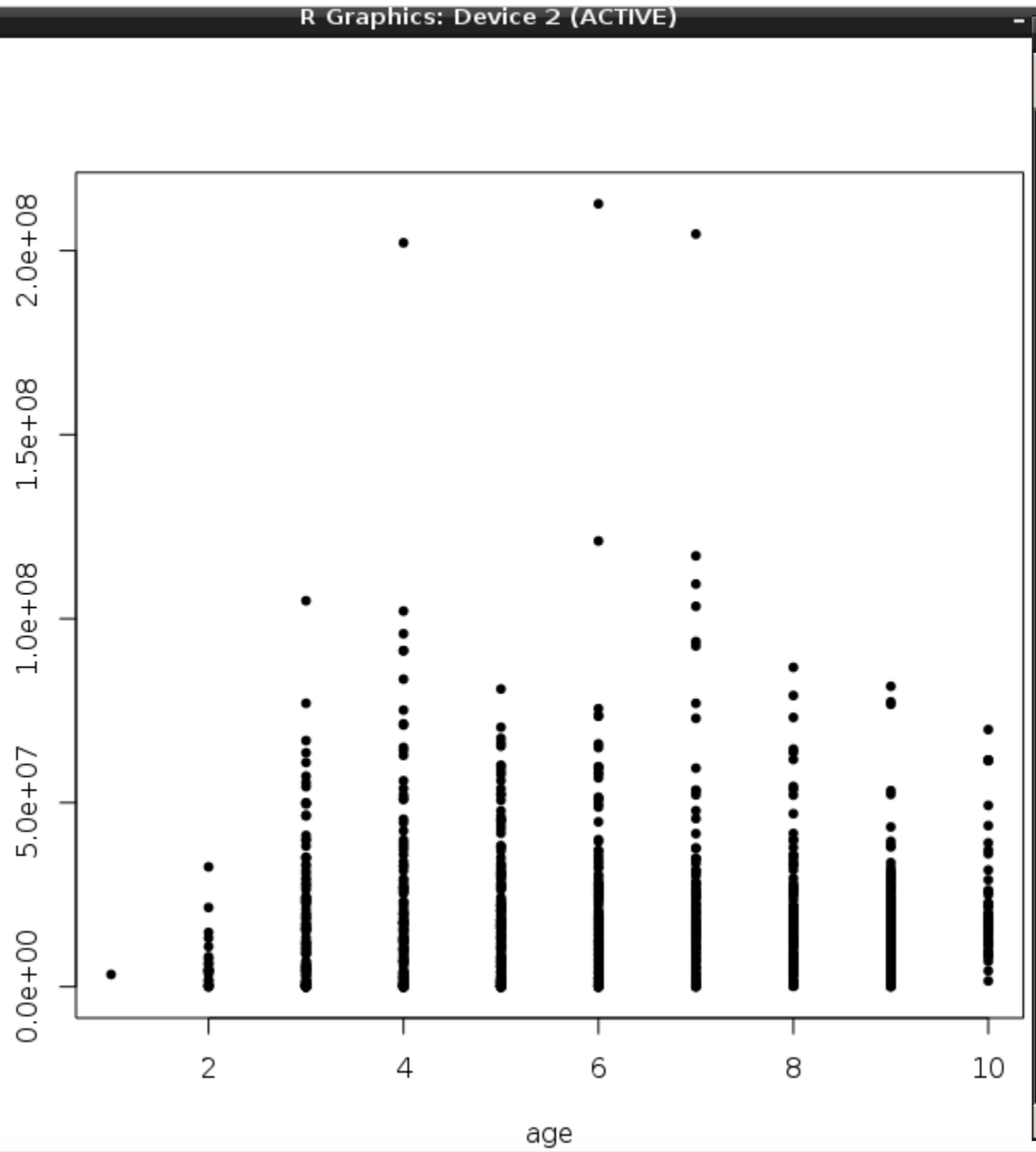
File Edit Tabs Help

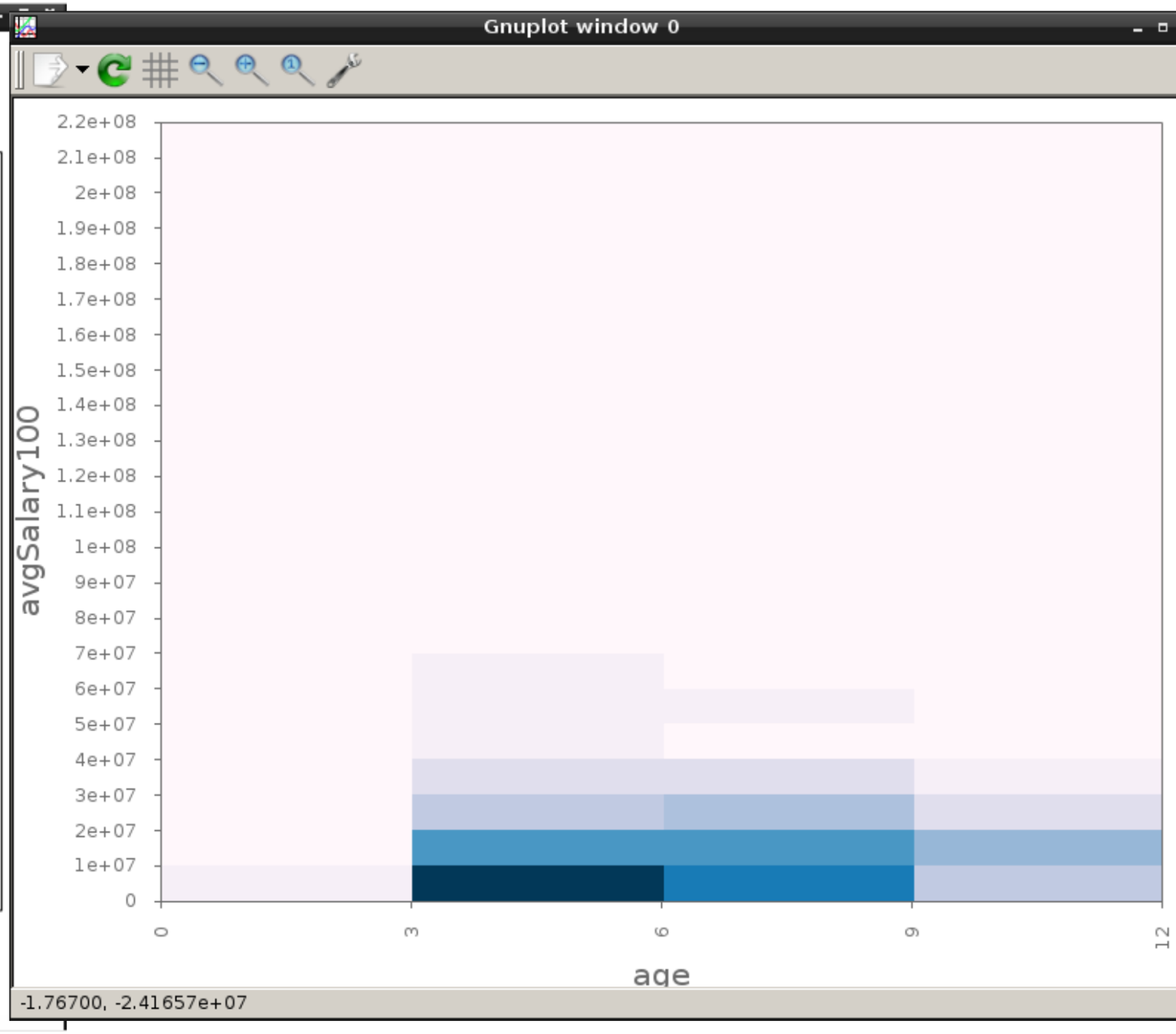
LXTerminal | LXTerminal

```
[sharemind@sm-build-vm rmind]$ ./rmind
Rmind
Copyright (C) Cybernetica AS
Type ‘q()’ to quit
Connecting to Sharemind...
Connected
> salary <- load("DS1", "salaries")
> subject <- load("DS1", "subjects")
> edu <- merge(subject, salary)
> age <- edu$age
> sal <- edu$avgSalary100
> heatmap (age, sal)
>
```

Rmind

The Rmind tool





Rmind

Features of Rmind

- **Data import**: CSV, anything with custom importers
- **Descriptive statistics:** stdev, var, cov, quantiles, histogram, frequency plots, heatmap
- **Quality assurance:** filtering, outlier removal with median absolute deviation
- **Transformations**: Sorting, merging, aggregation
- **Testing**: t-test, chi-square, Cochrane-Armitage, transmission disequilibrium, Wilcoxon, Mann-Whitney
- **Multiple testing:** Bonferroni correction, Benjamini-Hochberg procedure
- **Regressions**: linear, logistic
- We are continuously implementing new functions.

sharemind

Legal situation

- In January 2014, the Estonian Data Protection Agency cleared the use of Sharemind/Rmind for education records of Estonian students.
- In January 2015, the Estonian Tax and Customs Board cleared the use of Sharemind/Rmind for analyzing tax records of working students.
- We also have experience in forming contracts with all associated parties under European law.
- The EU PRACTICE project published a legal analysis of the technology from a European perspective. http://practice-project.eu/downloads/publications/D31.1-Risk-assessment-legal-status-PU-M12.pdf

sharemind

Literature

1. [K15] Liina Kamm. **Privacy-preserving statistical analysis using secure multi-party computation**. PhD thesis. University of Tartu. 2015. http://hdl.handle.net/10062/45343
2. [BKLS14] Dan Bogdanov, Liina Kamm, Sven Laur, Ville Sokk. **Rmind: a tool for cryptographically secure statistical analysis**. Cryptology ePrint Archive, Report 2014/512. 2014. http://eprint.iacr.org/2014/512.pdf
3. [KBLV13] Liina Kamm, Dan Bogdanov, Sven Laur, Jaak Vilo. **A new way to protect privacy in large-scale genome-wide association studies**. Bioinformatics 29 (7): 886-893, 2013. http://bioinformatics.oxfordjournals.org/content/29/7/886
4. [B13] Dan Bogdanov. **Sharemind: programmable secure computations with practical applications**. PhD thesis. University of Tartu. 2013. http://hdl.handle.net/10062/29041

sharemind

Acknowledgments

Our entry to the iDASH Privacy & Security Workshop Secure Genome Analysis Competition was prepared with support from

PRACTICE

http://practice-project.eu/

"The **PRACTICE** project has received funding from the European Union's Seventh Framework Programme ([FP7/2007-2013]) under grant agreement number ICT-609611."

sharemind