



EARL CONFERENCE 2018

EARL US Roadshow | November, 2018

Building an R-centered Reproducible Research Framework

Dr. Charles Tilford, Ph.D.

Translational Bioinformatics, BMS

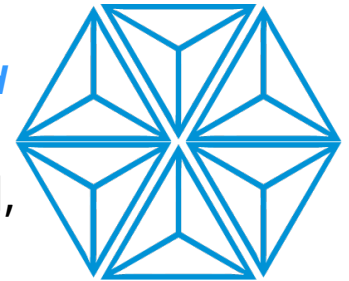
We are a Biopharma Company

Bristol-Myers Squibb

A global biopharmaceutical company whose mission is to discover, develop and deliver innovative medicines that help patients prevail over serious diseases.

Oncology, Immunoscience,
Cardiovascular, Fibrosis

Major research locations in NJ,
CA and (soon!) Cambridge, MA



- **Mechanism:**

How does this
drug work?



- **Personalized**

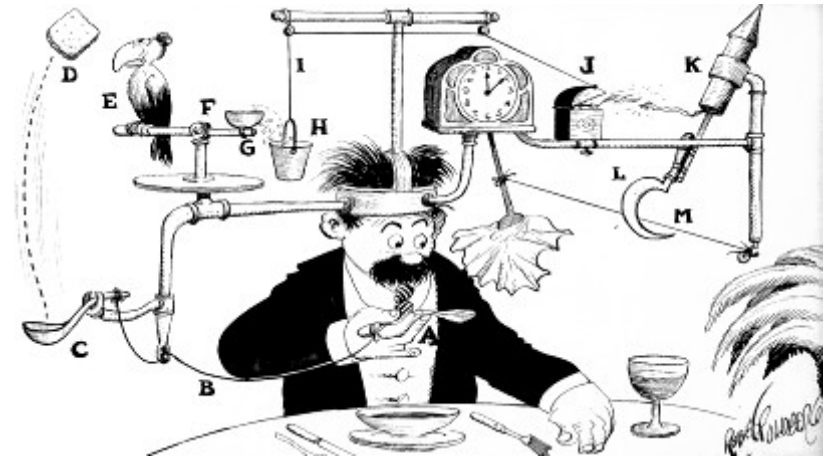
Medicine: Who's likely to
benefit from the drug?

- Can we find options for
those not benefiting?

translational
bioinformatics

- Provide answers with large
biological datasets
- Next Generation
Sequencing, proteomics, ...

Reproducible Research (RR) Improves the Way We Code



- Haphazardly constructed software hinders:
 - Updating or extending analyses
 - Merging or comparing results
 - Reproduction of results at a future date
- Our Reproducible Research framework allows:
 - Reliable access to previously used code and data
 - Improved documentation of analyses and workflows
 - Elimination of "Point & Click" tools and frameworks
 - Liberation of software from hardware






Using R for RR

aka the "Jolly Roger Approach"

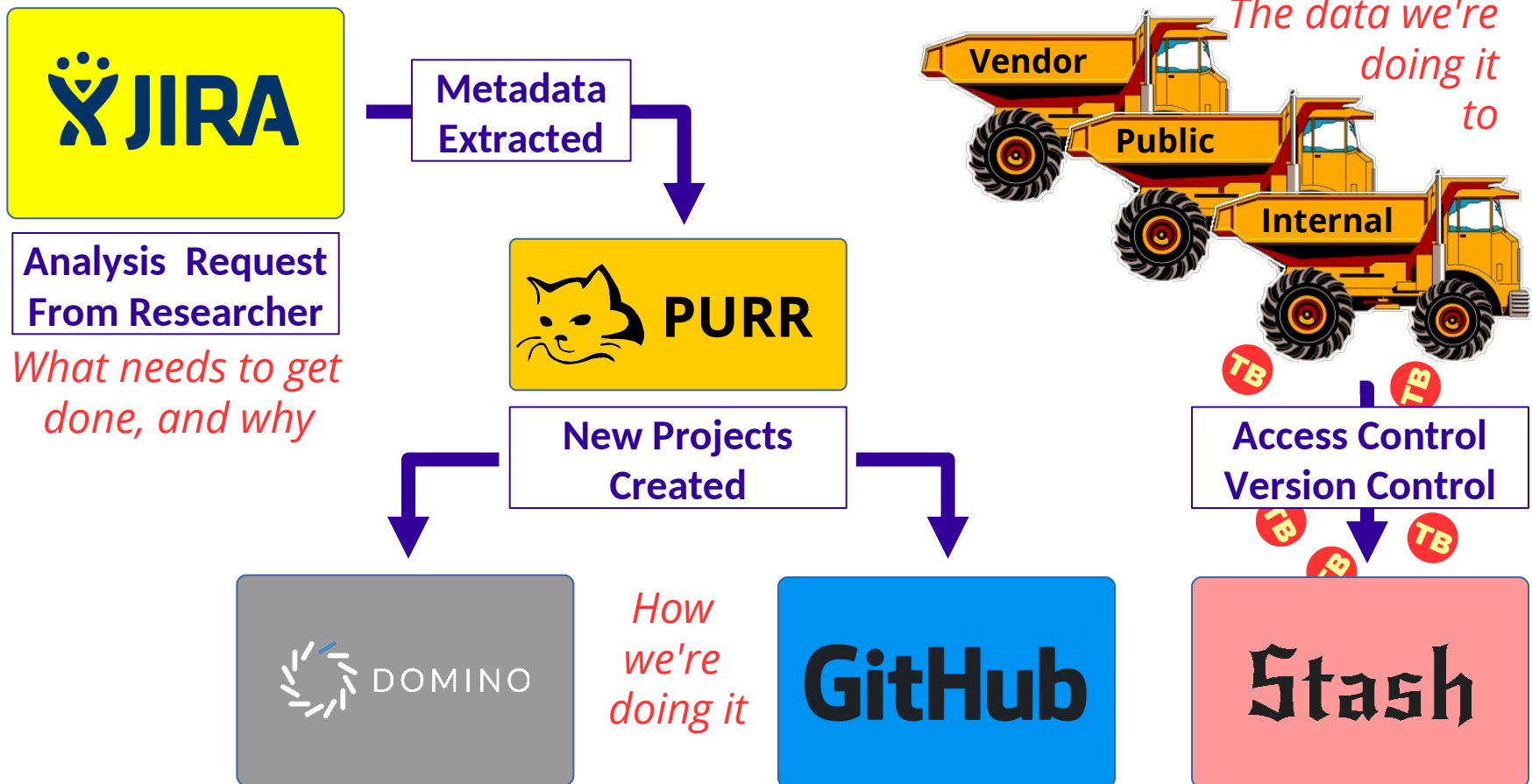
- R is an established analytical framework
- Open source, cross platform
 - BMS can easily contribute back to community
- Advanced, professional graphics via ggplot2
- Python / Matlab work ... but with R we also get:
 - **Bioconductor**, a massive bioinformatics package archive
 - **knitr**, links code and output via literate programming
 - **Rmarkdown**, gives us standardized knitr templates
 - **Shiny**, the ability to create interactive web reports

Framework components work together for reproducibility

BMS Tool 

Component	Tools	Philosophy
Initiation	Jira PURR 	Jira assures required metadata input, PURR initializes project assets following best practices
Data	Stash 	Versioned storage. Extended ACLs for strict access control, <code>sshfs</code> for secure identification.
Code	GitHub BRAN 	Project code in git, reusable modules in BRAN. Helper code to ease package development with Roxygen and testthat.
Analysis	Domino Docker	Virtualize away hardware. Develop as desired, but NO CLICKING in production - script only! Capturing "process" in GUI tools is untenable.
Reporting	knitr Shiny Domino	Rmarkdown templates for static reports via knitr. Shiny interactive exploration of huge result sets. Tag and version production result sets in Domino.

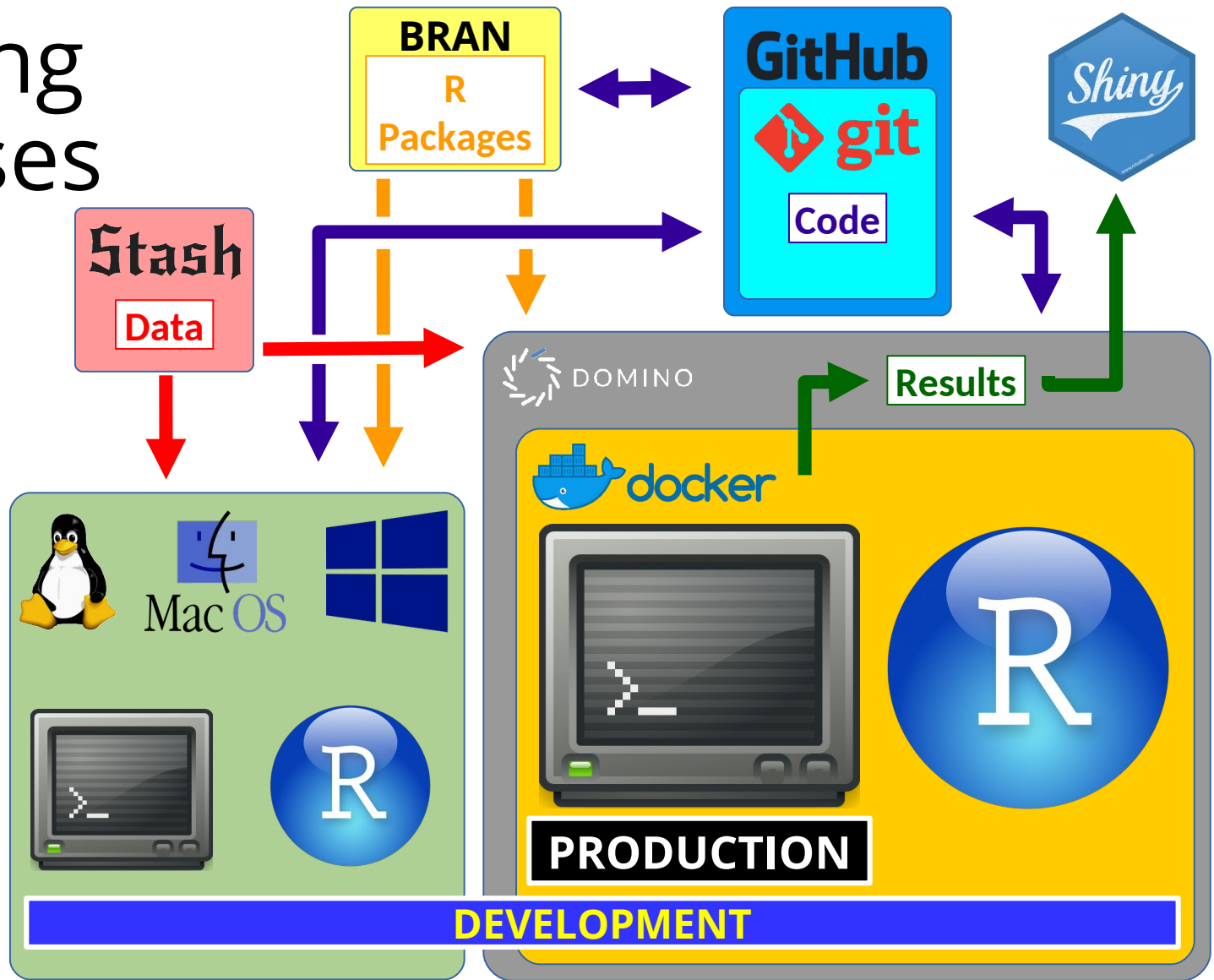
PURR & Stash- Laying the groundwork for an analysis



Running Analyses

Researchers are free to *develop* where and how they wish.

Production MUST be run in Domino, via script.



User-configured Reports via knitr

```
### Tumors Where Low r params$gene` Expression Associ

```{r hi_gene_low_hr, fig.height = 6, fig.width = 6}

pull out tumors where high gene ~ low hr
low_tumors <- filter(gep_v_os_bytumor, CI.r < 1)

arrange(low_tumors, desc(from1)) %>%
 select(-Tumor.Type, -from1) %>%
 kable(digits = 2)

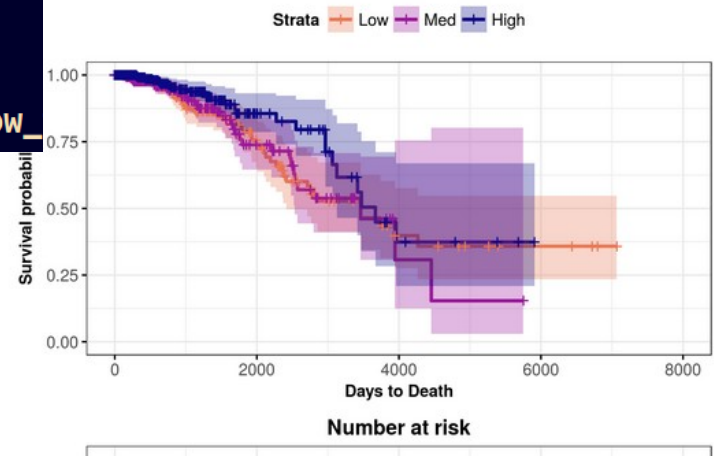
plot km for each tumor
low_tumor_data <- suppressWarnings(semi_join(gep, low
```

- Trivial to run new analysis
- Code and output are directly linked:
  - ... explicitly in Rmarkdown
  - ... by single click to show code in report

## 2.3.4 Tumors Where Low CD8A Expression Associates With Worse Prognosis

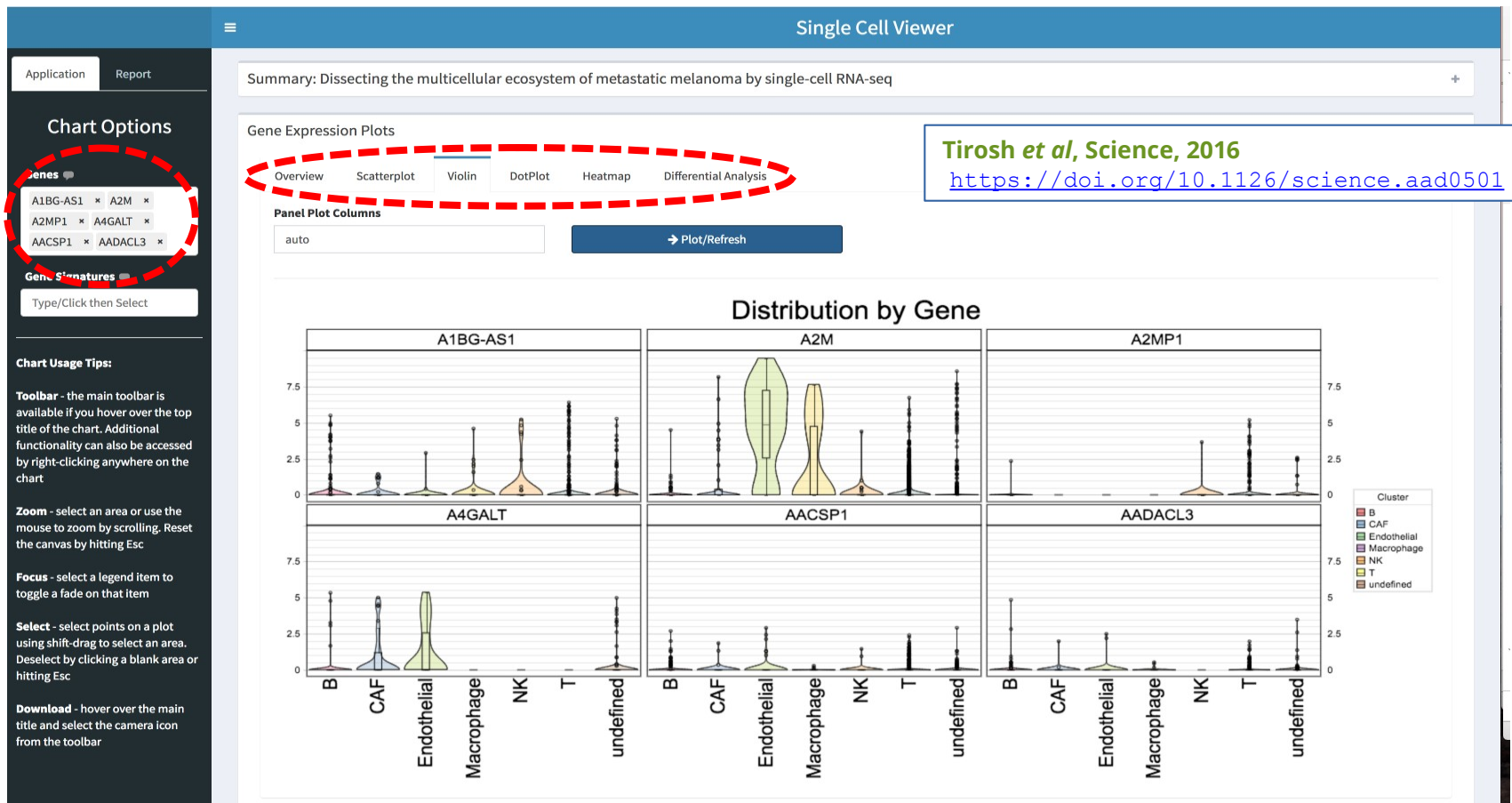
Tumor.Type	Disease	n events	HR	CI.l	CI.r	P.value
READ	Rectum adenocarcinoma	166	10	0.53	0.35 0.80	0.00
CESC	Cervical squamous cell carcinoma and endocervical adenocarcinoma	306	61	0.83	0.73 0.95	0.01
CM	Skin Cutaneous Melanoma	463	154	0.90	0.85 0.95	0.00
HC	Liver hepatocellular carcinoma	372	89	0.86	0.76 0.96	0.01
CA	Breast invasive carcinoma	1099	107	0.89	0.80 0.99	0.03

CD8A mRNA vs. Overall Survival in TCGA Breast invasive carcinoma





# Interactive Shiny Reports Allow Filtering and Exploration



# Culture - Getting the herd to drink



- Adopting the framework helps *you*
  - R is built for analysis
  - Bioconductor covers your typical needs, your colleagues likely have packages for others
  - RR makes it easy to revise and update your code
  - Collaborators more likely to respect your findings
    - Results can be easily shared and explored
- Implementation is guided by internal discussion and grass-roots initiatives
  - Adoption is an expectation, not a suggestion

# Summary

- Platform deployed to 50+ coders & analysts
  - Used daily. Proven utility in collaboration, code reuse, project hand-off, rapid reanalysis
  - Biologists, statisticians, programmers, engineers
- R is well suited for pharmaceutical research
  - Bioconductor, Rmd, knitr and Shiny jump-started RR
- Continuous development, currently working on:
  - Results dissemination
  - Control of multiple production runs

# Acknowledgements

## Authors

Scott Chasalow  
Don Jackson (Rmd templates)  
Isaac Neuhaus  
Mark Russo (PURR)  
Charles Tilford (BRAN)

## Coders

Matt Ralston (Stash)  
Andrew Smith (Stash)  
John Thompson (Rmd)  
Ron Ammar (Rmd)  
Connie Brett (Shiny Apps)

## Partners

Aggregate Genius  
Domino (Mark Silverberg & Team)

## Tools

RStudio  
Jira  
GitHub

## Leadership

Joe Szustakowski

# External Media Sources

- BMS Logo - [bms.com](https://bms.com)
- Domino Logo - [Domino Blog](https://domino.blog)
- R Logo - [Wikipedia](https://en.wikipedia.org/wiki/R_(programming_language))
- Docker Logo - [Docker](https://www.docker.com)
- Windows Logo - [Wikimedia](https://en.wikipedia.org/wiki/Microsoft_Windows)
- MacOS Logo - [Wikipedia](https://en.wikipedia.org/wiki/MacOS)
- Linux Tux - [Wikipedia](https://en.wikipedia.org/wiki/Tux_(Linux))
- GitHub Logo - [Wikimedia](https://en.wikipedia.org/wiki/GitHub)
- git Logo - [Wikimedia](https://en.wikipedia.org/wiki/Git)
- JIRA Logo - [Atlassian](https://atlassian.com)
- Shiny Badge - [rstudio.com](https://www.rstudio.com)
- RStudio Logo - [rstudio.com](https://www.rstudio.com)
- RStudio Image - [Wikipedia](https://en.wikipedia.org/wiki/R_(programming_language))
- Rube Goldberg - [Wikimedia](https://en.wikipedia.org/wiki/Rube_Goldberg)
- Water Hole - [Wikimedia](https://en.wikipedia.org/wiki/Water_Hole)
- GTK Widgets - [gtk.org](https://gtk.org)
- "Prohibited" - [Wikimedia](https://en.wikipedia.org/wiki/No_entry_sign)
- Jolly Roger - [Wikimedia](https://en.wikipedia.org/wiki/Jolly_Roger)
- Terminal Icon - [Wikimedia](https://en.wikipedia.org/wiki/Terminal_(computer_science))
- Dump Truck - [Wikimedia](https://en.wikipedia.org/wiki/Dump_truck)