

IMEter v2.1

UCDAVIS GENOME CENTER

- * A new version of the IMEter now predicts the expression-increasing ability of introns in over 30 different plant species
- * Adapting a bioinformatics tool from one, well annotated, genome to many genomes of varying quality is not straightforward
- * Just because you *can* use a newly sequenced genome, it doesn't always mean that you should



Keith Bradnam



Ian Korf*



Alan Rose*

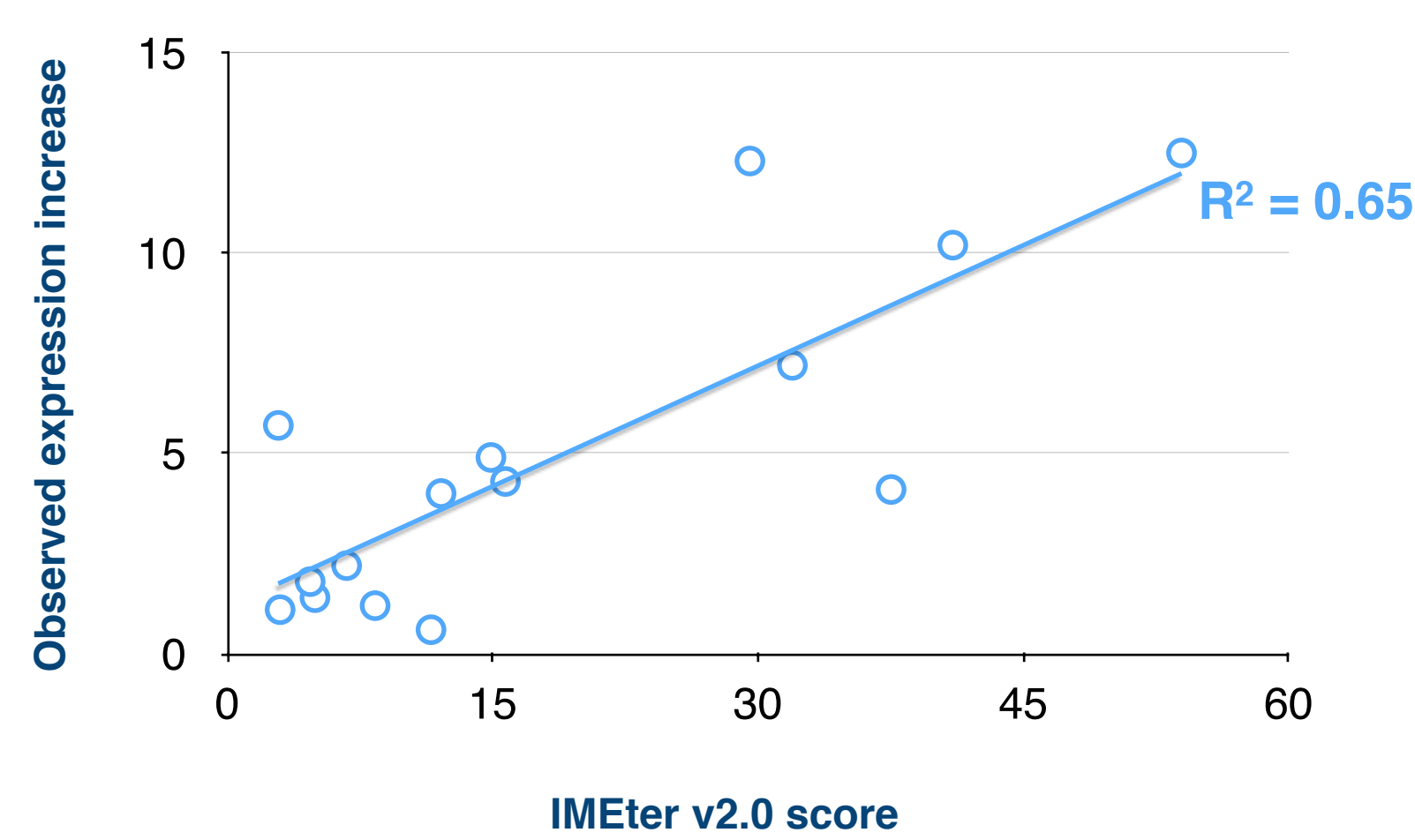
BACKGROUND



Image from <http://biosci.unl.edu/>

- * Some introns can enhance gene expression by a process called intron-mediated enhancement (IME)
- * In *Arabidopsis thaliana*, enhancing introns can increase levels of mRNA and protein by as much as 13-fold
- * Enhancing introns tend to be close to the transcription start site (typically 1st introns)
- * It can take up to **one year** to experimentally test wild-type introns in *A. thaliana*
- * We previously developed a software tool, the IMEter, that can predict the enhancing ability of any *A. thaliana* intron in **seconds**

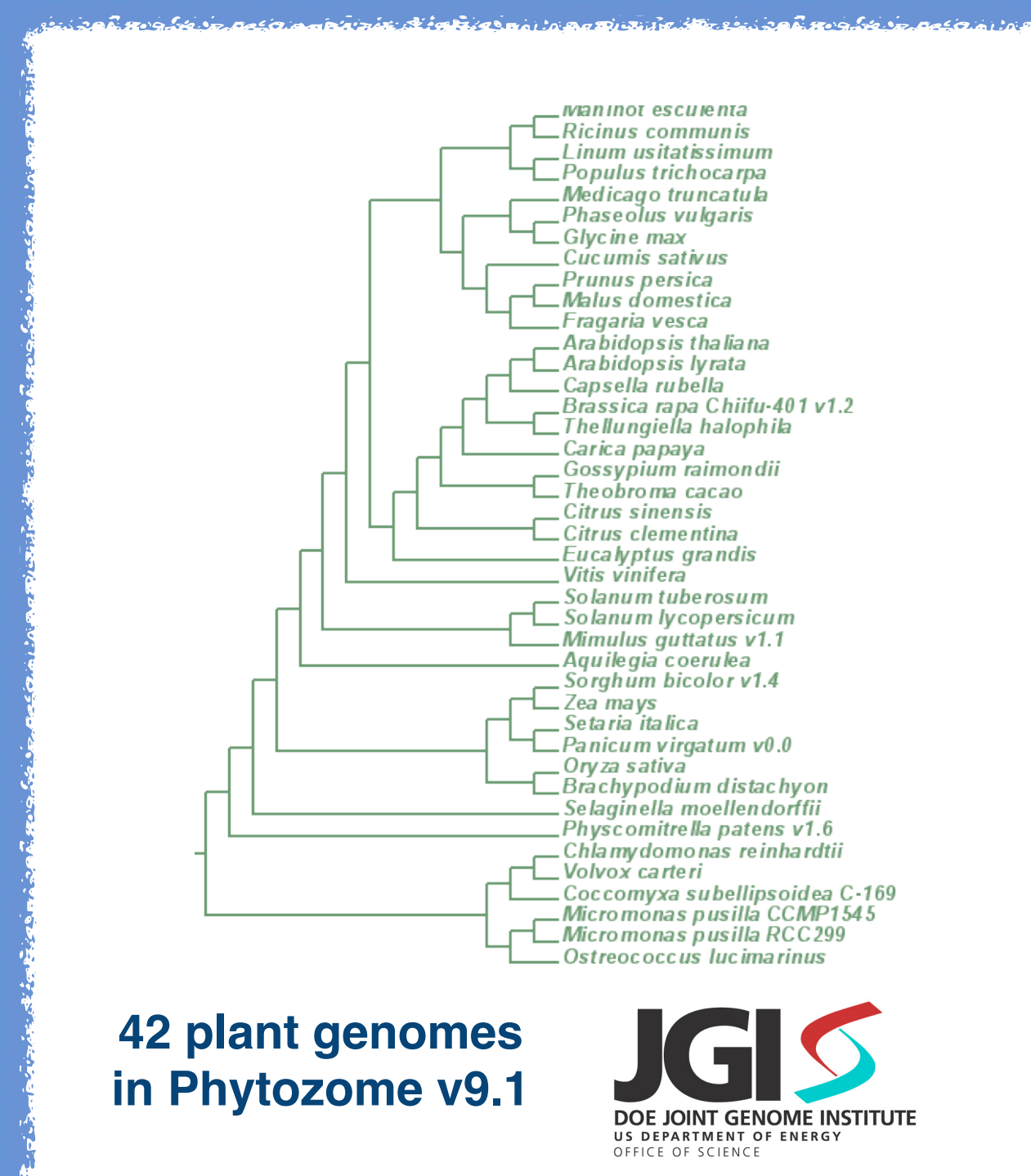
The IMEter predicts how well introns enhance expression



THE PROBLEM

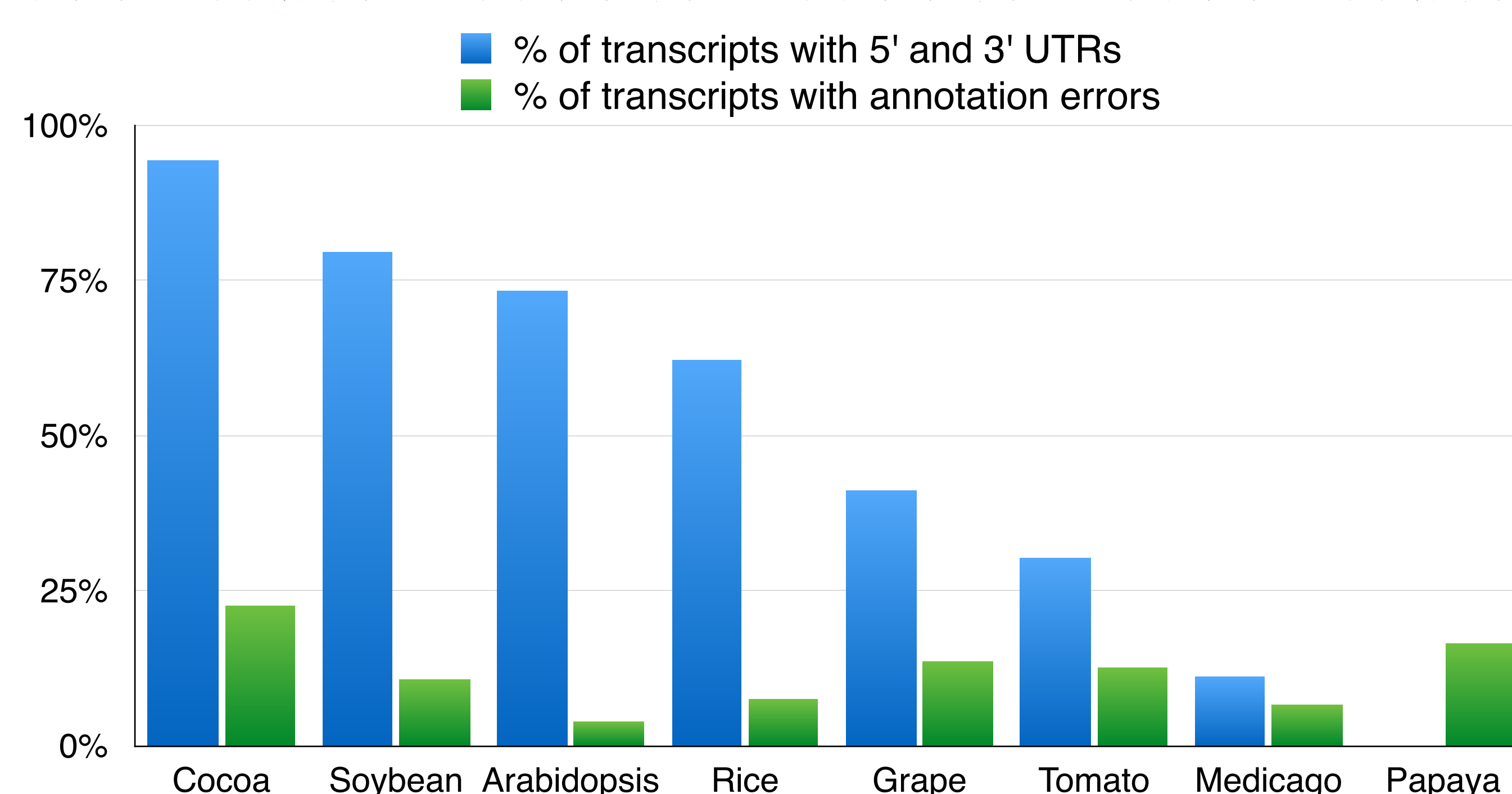
- * IMEter v2.0 has only been trained in *A. thaliana* and the code relies on some information that is specific to The Arabidopsis Information Resource (TAIR)
- * There are now lots of published plant genomes and many researchers may wish to know whether a particular intron in rice, grape, corn etc. might enhance gene expression

USING PHYTOZOME



- * The Phytozome database contains genome sequences and annotations for many plants
- * Even if the species were not sequenced by the JGI, they still provide a consistent organization of files and identifiers
- * This consistency means that bioinformaticians don't have to smash their heads against their computers in frustration
- * This enables us to more easily train the IMEter in many different species (even when we only have experimental IME data for *A. thaliana*)

VARIATION IN ANNOTATION QUALITY



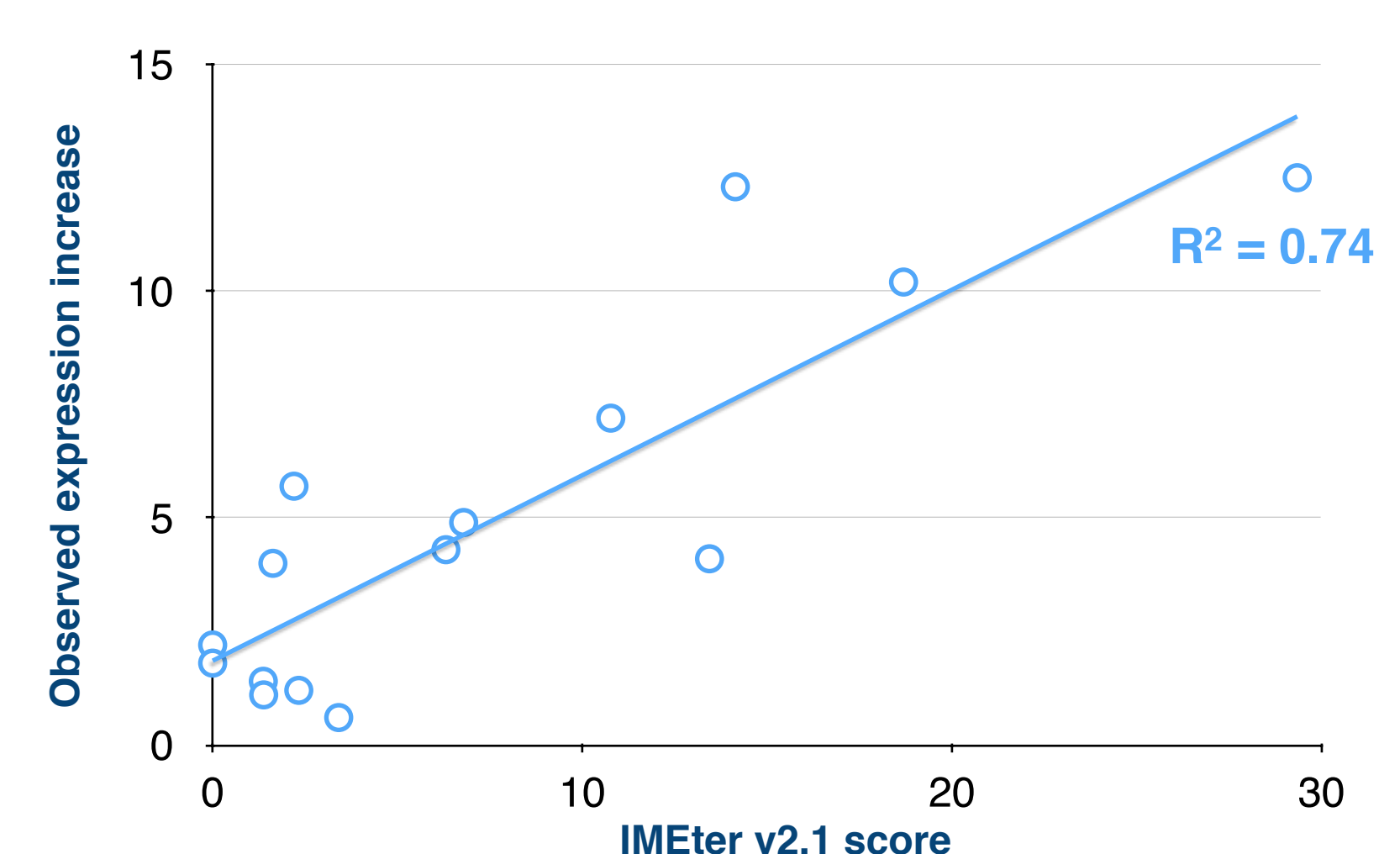
- * 5' UTR annotations are important because many UTRs contain introns, the position of which is used in IMEter training
- * Annotation errors include internal stop codons, non-canonical splice sites, missing start and/or stop codons, etc.
- * We ignore any transcript with annotation errors
- * The strawberry genome (*Fragaria vesca*) uses 276 different splice donor/acceptor combinations!

Top 10 splice site combinations in strawberry genome

Donor site	Acceptor site	Frequency
GT	AG	131,651
NN	NN	63
TT	AA	37
TT	TT	36
AA	TT	35
AT	AA	34
AA	AA	32
TC	TT	30
TT	CA	28
AT	TT	26

IMEter v2.1

New IMEter is a better predictor of expression



- * The new IMEter can predict the enhancing ability of introns in 34 different plant species
- * Several species in Phytozome could not be used due to poor annotation quality
- * Other improvements to the IMEter code means that it now is an even better predictor of expression (in *A. thaliana*, R^2 increased from 0.65 to 0.74)
- * Further improvements are planned for a v2.2 release which will also leverage data from an improved version of Phytozome (v10)

**NOT ALL GENOMES
ARE ANNOTATED EQUALLY!**

THE SMALL PRINT

Alan Rose and Ian Korf hold appointments in the Department of Molecular and Cellular Biology. The IMEter (v2.0) was published in Nucleic Acids Research (2011). A manuscript for IMEter v2.2 is being prepared. Your home is at risk if you do not keep up with mortgage repayments.



<http://imeter.korflab.com>