Report writing

Letter 1: Marguerat 2012

title

Quantitative Analysis of Fission Yeast Transcriptomes and Proteomes in Proliferating and Quiescent Cells

Does the title give you information?
What does it tell you?

Letter 1: Marguerat 2012

Summary (same as abstract)

Data on absolute molecule numbers will empower the modeling, understanding, and comparison of cellular functions and biological systems.

What does this part do?

We quantified transcriptomes and proteomes in fission yeast during cellular proliferation and quiescence. This rich resource provides the first comprehensive reference for all RNA and most protein concentrations in a eukaryote under two key physiological conditions. The integrated data set supports quantitative biology and affords unique insights into cell regulation.

What does this part do?

Although mRNAs are typically expressed in a narrow range above 1 copy/cell, most long, noncoding RNAs, except for a distinct subset, are tightly repressed below 1 copy/cell. Cell-cycle-regulated transcription tunes mRNA numbers to phase-specific requirements but can also bring about more switch-like expression. Proteins greatly exceed mRNAs in abundance and dynamic range, and concentrations are regulated to functional demands. Upon transition to quiescence, the proteome changes substantially, but, in stark contrast to mRNAs, proteins do not uniformly decrease but scale with cell volume.

What does this part do?

Letter 1: Marguerat 2012

introduction

How do the authors mark sections of the introduction?

SUMMARY

Data on absolute molecule numbers will empower the modeling, understanding, and comparison of cellular functions and biological systems. We quantified transcriptomes and proteomes in fission yeast during cellular proliferation and quiescence. This rich resource provides the first comprehensive reference for all RNA and most protein concentrations in a eukaryote under two key physiological conditions. The integrated data set supports quantitative biology and affords unique insights into cell regulation. Although mRNAs are typically expressed in a narrow range above 1 copy/cell, most long, noncoding RNAs, except for a distinct subset, are tightly repressed below 1 copy/cell. Cell-cycle-regulated transcription tunes mRNA numbers to phase-specific requirements but can also bring about more switch-like expression. Proteins greatly exceed mRNAs in abundance and dynamic range, and concentrations are regulated to functional demands. Upon transition to quiescence, the proteome changes substantially, but, in stark contrast to mRNAs, proteins do not uniformly decrease but scale with cell volume.

INTRODUCTION

Gene regulation is crucial to implement genomic information and to shape properties of cells and organisms. Transcriptomes and proteomes are dynamically tuned to the requirements of cell volume, physiology and external factors. Although transcriptomic and proteomic approaches have provided ample data on relative expression changes between different conditions, little is known about actual numbers of RNAs and proteins within cells and how gene regulation affects these numbers. More generally, most data in biology are qualitative or relatively quantitative, but ultimately many biological processes will only be understood if investigated with absolute quantitative data to support mathematical modeling. Other areas of science have long appreciated the limits of relative, or compositional, data and potential pitfalls of their naive analysis (Lovell et al., 2011).

Insights into numbers and cell-to-cell variability of selected mRNAs and proteins have been provided by single-cell studies (Larson et al., 2009), but these approaches require genetic manipulation and are not well suited for genome-scale analyses. Relating mRNA to protein abundance in single cells is challenging, with only one such study available for a prokaryote (Taniguchi et al., 2010). Global mRNA abundance for yeast populations have been estimated (Holstege et al., 1998; Miura et al., 2008). There are no comparisons for cellular concentrations of mRNAs and the emerging diversity of noncoding RNAs.

RNA-seq now allows actual counting of RNA numbers, offering unbiased genome-wide information on average cellular RNA concentrations in cell populations (Ozsolak and Milos, 2011). Moreover, the global quantification of proteins has recently become possible owing to advances in mass spectrometry, giving valuable insight into the protein content of different cells (Beck et al., 2011; Cox and Mann, 2011; Maier et al., 2011; Nagaraj et al., 2011; Vogel and Marcotte, 2012).

Here, we combine quantitative RNA-seq and mass spectrometry to analyze at unprecedented detail and scale how changes in cell physiology and volume are reflected in the cellular concentrations of all coding and noncoding RNAs and most proteins. We analyze two fundamental physiological states in fission yeast: (1) proliferating cells that need to constantly replenish their RNAs and proteins, and (2) postmitotic cells that do not grow or divide owing to nitrogen limitation and reversibly arrest in a quiescent state (Yanagida, 2009). Although quiescent states are common, both for yeast and for cells in the human body, most research has focused on proliferating cells. The ability to alternate between proliferation and quiescence is central to tissue homeostasis and renewal, pathophysiology, and the response to life-threatening challenges (Coller, 2011). For example, quiescent lymphocytes

How is this paragraph of the introduction different?

Letter 1: Marguerat 2012

Figure 1: your headline shot

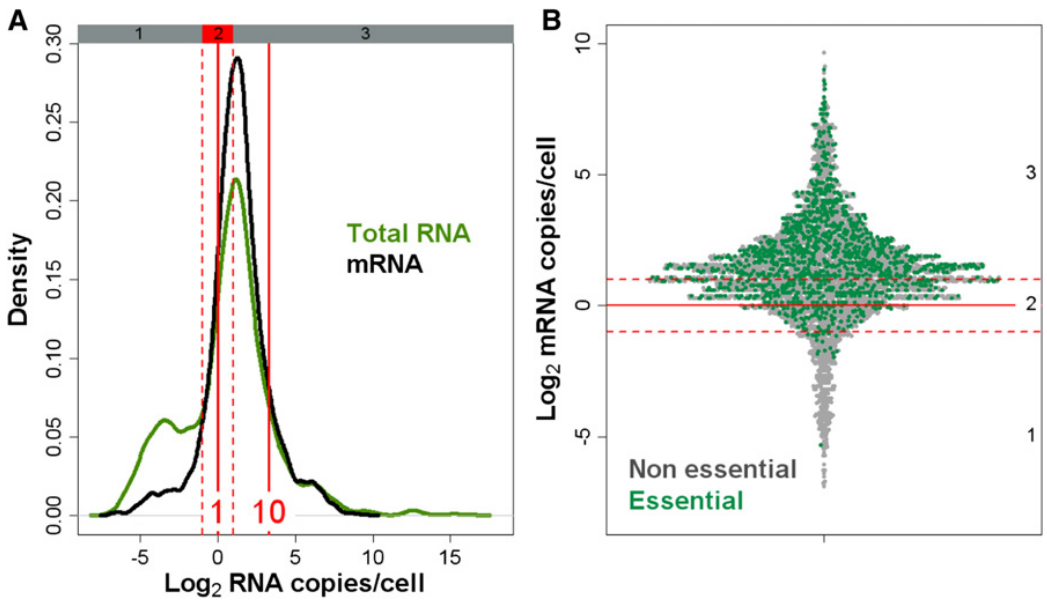

Figure 1. Transcriptome Quantification in Proliferating Cells

(A) Abundance distribution of total RNA (green) and mRNA (black). Red vertical lines indicate 1 and 10 RNA copies/cell, and red hatched lines delimit expression zones 1 to 3. See also Figure S1 and Table S10.

(B) Abundance for all detected mRNAs (each dot represents a gene). Green and gray dots correspond to essential and non essential genes, respectively. Expression zones are indicated at right.

Look at part B.
How complex is it?
How would you rate its *explanatory power*?

Letter 1: Marguerat 2012

Results and discussion

Note that they use subheadings.

Note how one of them gives you the main result in <u>one sentence.</u>

and dermal fibroblasts become activated to mount immune responses or support wound healing, respectively. Adult stem cells also alternate between proliferating and quiescent states, and the deregulation of either state can cause complex pathologies such as cancer (Li and Clevers, 2010).

Our integrated transcriptomic and proteomic data, acquired in parallel under highly controlled conditions in a simple model, afford varied biological insights and reveal key principles of RNA and protein expression in proliferating and quiescent cells with broad relevance for other eukaryotes. This rich resource also provides a quantitative framework toward a systems-level understanding of genome regulation, and the common units of the absolute data allow direct comparison of different biological processes and organisms.

RESULTS AND DISCUSSION

Transcriptome and Proteome Quantification in Two Conditions

We acquired quantitative expression data relative to absolutely calibrated standards for transcriptomes and proteomes of haploid fission yeast cells. For transcripts, genome-wide measurements were obtained by calibrating RNA-seq data from total RNA preparations with data on absolute cellular concentrations for 49 mRNAs, covering the dynamic expression range. The overall measurement error was estimated to be ~2-fold or less (Figure S1; Tables S1–S4 available online). Protein quantification was performed on the same cell samples using a mass spectrometry (MS) approach (Schmidt et al., 2011). Selected proteotypic peptides from 39 proteins (Table S5), covering the dynamic expression range, were used to absolutely quantify the corresponding proteins (Tables S6 and S7). These data were then used to translate the MS-intensities for the other proteins into estimates of cellular concentration (Figures S2A–S2D and S3; and Tables S8 and S9). The mean overall measurement error was estimated at 2.4- and 2.7-fold for proliferating and quiescent cells, respectively.

We quantified transcriptomes and proteomes in two distinct physiological conditions: (1) exponentially proliferating cells in quiescent cells. Table S4 provides the cellular copy numbers for RNAs and proteins in the two conditions.

Most mRNAs Are Expressed in Narrow Range above 1 Copy/Cell

In proliferating cells, we measured a total of ~41,000 mRNA molecules/cell on average, representing ~5% of the overall ~802,000 rRNAs/cell in our samples. Protein-coding genes produced a median of 2.4 mRNA copies/cell, ranging from ~0.01 to >810 copies (Figure 1A). Only 71 genes showed no detectable mRNA signal, 43 of which are annotated as "dubious" or "orphan" (Wood et al., 2012). To discuss our findings, we distinguished three somewhat arbitrary expression zones, set relative to the one RNA copy/cell mark (Figure 1A). Zone 1 contained low-abundance mRNAs detected at <0.5 copies/cell. Zone 2 mRNAs were expressed at ~1 copy/cell (0.5–2 copies), where fluctuations due to cell division or stochastic expression will strongly affect the presence of mRNAs in cells. Zone 3 mRNAs showed more robust expression at >2 copies/cell. Most mRNAs were expressed within a low and narrow range: whereas >90% of all annotated mRNAs (4,608/5,110) belonged to zones 2 or 3, 86.1% of these mRNAs were present at <10 copies/cell (Figure 1A). Low overall mRNA concentrations have also been reported for budding yeast, which has comparable gene numbers and cell size, with even lower estimates for median mRNA abundance (<1 copy/cell) and total mRNA molecules/cell (Holstege et al., 1998; Miura et al., 2008). Our findings are in line with a single-cell study of budding yeast, where five mRNAs show 2.6–13.4 copies/cell, with a total estimate of 60,000 mRNA molecules/cell (Zenklusen et al., 2008).

We examined the mRNAs of the 1,273 genes essential for growth (Kim et al., 2010), which are expected to be expressed in proliferating cells. Nearly all essential mRNAs were expressed in zones 2 or 3 (98.4%; Figure 1B). This finding raises the possibility that ~1 mRNA copy/cell defines a natural minimal threshold for productive gene expression.

The view of ~1 mRNA copy/cell as an expression threshold is supported by recent data from metazoa, where mRNA levels

Letter 1: Marguerat 2012

Results and discussion

and dermal fibroblasts become activated to mount immune responses or support wound healing, respectively. Adult stem cells also alternate between proliferating and quiescent states, and the deregulation of either state can cause complex pathologies such as cancer (Li and Clevers, 2010).

Our integrated transcriptomic and proteomic data, acquired in parallel under highly controlled conditions in a simple model, afford varied biological insights and reveal key principles of RNA and protein expression in proliferating and quiescent cells with broad relevance for other eukaryotes. This rich resource also provides a quantitative framework toward a systems-level understanding of genome regulation, and the common units of the absolute data allow direct comparison of different biological processes and organisms.

RESULTS AND DISCUSSION

Transcriptome and Proteome Quantification in Two Conditions

We acquired quantitative expression data relative to absolutely calibrated standards for transcriptomes and proteomes of haploid fission yeast cells. For transcripts, genome-wide measurements were obtained by calibrating RNA-seq data from total RNA preparations with data on absolute cellular concentrations for 49 mRNAs, covering the dynamic expression range. The overall measurement error was estimated to be ~2-fold or less (Figure S1; Tables S1–S4 available online). Protein quantification was performed on the same cell samples using a mass spectrometry (MS) approach (Schmidt et al., 2011). Selected proteotypic peptides from 39 proteins (Table S5), covering the dynamic expression range, were used to absolutely quantify the corresponding proteins (Tables S6 and S7). These data were then used to translate the MS-intensities for the other proteins into estimates of cellular concentration (Figures S2A–S2D and S3; and Tables S8 and S9). The mean overall measurement error was estimated at 2.4- and 2.7-fold for proliferating and quiescent cells, respectively.

We quantified transcriptomes and proteomes in two distinct physiological conditions: (1) exponentially proliferating cells in quiescent cells. Table S4 provides the cellular copy numbers for RNAs and proteins in the two conditions.

Most mRNAs Are Expressed in Narrow Range above 1 Copy/Cell

In proliferating cells, we measured a total of ~41,000 mRNA molecules/cell on average, representing ~5% of the overall ~802,000 rRNAs/cell in our samples. Protein-coding genes produced a median of 2.4 mRNA copies/cell, ranging from ~0.01 to >810 copies (Figure 1A). Only 71 genes showed no detectable mRNA signal, 43 of which are annotated as "dubious" or "orphan" (Wood et al., 2012). To discuss our findings, we distinguished three somewhat arbitrary expression zones, set relative to the one RNA copy/cell mark (Figure 1A). Zone 1 contained low-abundance mRNAs detected at <0.5 copies/cell. Zone 2 mRNAs were expressed at ~1 copy/cell (0.5–2 copies), where fluctuations due to cell division or stochastic expression will strongly affect the presence of mRNAs in cells. Zone 3 mRNAs showed more robust expression at >2 copies/cell. Most mRNAs were expressed within a low and narrow range: whereas >90% of all annotated mRNAs (4,608/5,110) belonged to zones 2 or 3, 86.1% of these mRNAs were present at <10 copies/cell (Figure 1A). Low overall mRNA concentrations have also been reported for budding yeast, which has comparable gene numbers and cell size, with even lower estimates for median mRNA abundance (<1 copy/cell) and total mRNA molecules/cell (Holstege et al., 1998; Miura et al., 2008). Our findings are in line with a single-cell study of budding yeast, where five mRNAs show 2.6–13.4 copies/cell, with a total estimate of 60,000 mRNA molecules/cell (Zenklusen et al., 2008).

We examined the mRNAs of the 1,273 genes essential for growth (Kim et al., 2010), which are expected to be expressed in proliferating cells. Nearly all essential mRNAs were expressed in zones 2 or 3 (98.4%; Figure 1B). This finding raises the possibility that ~1 mRNA copy/cell defines a natural minimal threshold for productive gene expression.

The view of ~1 mRNA copy/cell as an expression threshold is supported by recent data from metazoa, where mRNA levels

Where are?

- **Results**
- **Discussion**
- **References to other studies**

Letter 1: Marguerat 2012

Results and discussion

and dermal fibroblasts become activated to mount immune responses or support wound healing, respectively. Adult stem cells also alternate between proliferating and quiescent states, and the deregulation of either state can cause complex pathologies such as cancer (Li and Clevers, 2010).

Our integrated transcriptomic and proteomic data, acquired in parallel under highly controlled conditions in a simple model, afford varied biological insights and reveal key principles of RNA and protein expression in proliferating and quiescent cells with broad relevance for other eukaryotes. This rich resource also provides a quantitative framework toward a systems-level understanding of genome regulation, and the common units of the absolute data allow direct comparison of different biological processes and organisms.

RESULTS AND DISCUSSION

Transcriptome and Proteome Quantification in Two Conditions

We acquired quantitative expression data relative to absolutely calibrated standards for transcriptomes and proteomes of haploid fission yeast cells. For transcripts, genome-wide measurements were obtained by calibrating RNA-seq data from total RNA preparations with data on absolute cellular concentrations for 49 mRNAs, covering the dynamic expression range. The overall measurement error was estimated to be ~2-fold or less (Figure S1; Tables S1–S4 available online). Protein quantification was performed on the same cell samples using a mass spectrometry (MS) approach (Schmidt et al., 2011). Selected proteotypic peptides from 39 proteins (Table S5), covering the dynamic expression range, were used to absolutely quantify the corresponding proteins (Tables S6 and S7). These data were then used to translate the MS-intensities for the other proteins into estimates of cellular concentration (Figures S2A–S2D and S3; and Tables S8 and S9). The mean overall measurement error was estimated at 2.4- and 2.7-fold for proliferating and quiescent cells, respectively.

We quantified transcriptomes and proteomes in two distinct physiological conditions: (1) exponentially proliferating cells in quiescent cells. Table S4 provides the cellular copy numbers for RNAs and proteins in the two conditions.

Most mRNAs Are Expressed in Narrow Range above 1 Copy/Cell

In proliferating cells, we measured a total of ~41,000 mRNA molecules/cell on average, representing ~5% of the overall ~802,000 rRNAs/cell in our samples. Protein-coding genes produced a median of 2.4 mRNA copies/cell, ranging from ~0.01 to >810 copies (Figure 1A). Only 71 genes showed no detectable mRNA signal, 43 of which are annotated as "dubious" or "orphan" (Wood et al., 2012). To discuss our findings, we distinguished three somewhat arbitrary expression zones, set relative to the one RNA copy/cell mark (Figure 1A). Zone 1 contained low-abundance mRNAs detected at <0.5 copies/cell. Zone 2 mRNAs were expressed at ~1 copy/cell (0.5–2 copies), where fluctuations due to cell division or stochastic expression will strongly affect the presence of mRNAs in cells. Zone 3 mRNAs showed more robust expression at >2 copies/cell. Most mRNAs were expressed within a low and narrow range: whereas >90% of all annotated mRNAs (4,608/5,110) belonged to zones 2 or 3, 86.1% of these mRNAs were present at <10 copies/cell (Figure 1A). Low overall mRNA concentrations have also been reported for budding yeast, which has comparable gene numbers and cell size, with even lower estimates for median mRNA abundance (<1 copy/cell) and total mRNA molecules/cell (Holstege et al., 1998; Miura et al., 2008). Our findings are in line with a single-cell study of budding yeast, where five mRNAs show 2.6–13.4 copies/cell, with a total estimate of 60,000 mRNA molecules/cell (Zenklusen et al., 2008).

We examined the mRNAs of the 1,273 genes essential for growth (Kim et al., 2010), which are expected to be expressed in proliferating cells. Nearly all essential mRNAs were expressed in zones 2 or 3 (98.4%; Figure 1B). This finding raises the possibility that ~1 mRNA copy/cell defines a natural minimal threshold for productive gene expression.

The view of ~1 mRNA copy/cell as an expression threshold is supported by recent data from metazoa, where mRNA levels

results

discussion

referencing another study

Letter 1: Marguerat 2012

Figure 3

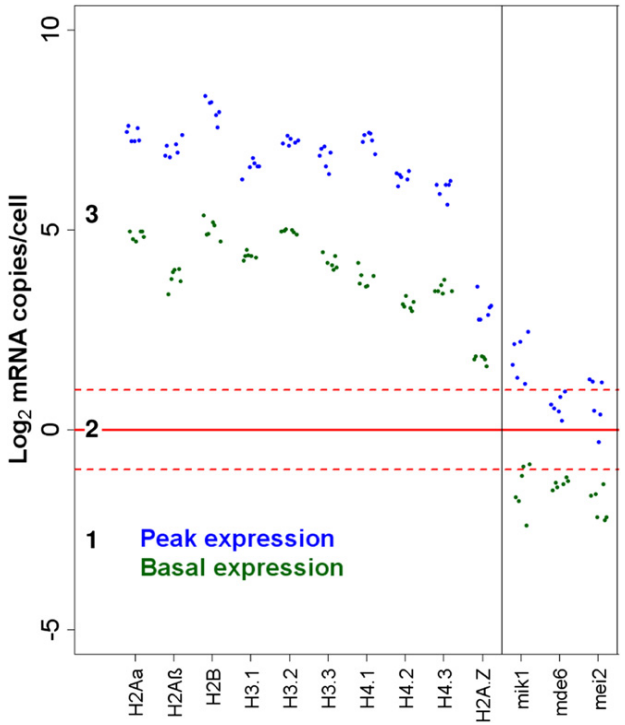

Figure 3. mRNA Copy Number Changes during Cell Cycle
Peak (blue) and basal (green) mRNA abundance of cell-cycle-regulated genes extrapolated from average data in asynchronous cultures, with 10% of cell-cycle assumed as duration for peak expression. Data for six cell-cycle time course experiments are indicated by clustered dots (Rustici et al., 2004). Left: ten histone mRNAs peaking during S phase; right: *mik1*, *mde6*, and *mei2* mRNAs peaking during M and G1 phases.
See also Figure S5 and Table S12.

First look at this figure without the figure legend.

Do you get the main conclusion?

Is this plot clear?
Why?

Letter2: Miller 2016

title

Miller *et al. BMC Genomics* (2016) 17:500
DOI 10.1186/s12864-016-2775-2

BMC Genomics

RESEARCH ARTICLE Open Access

Elucidation of the genetic basis of variation for stem strength characteristics in bread wheat by Associative Transcriptomics

Charlotte N. Miller[1†], Andrea L. Harper[1,4†], Martin Trick[1], Peter Werner[2], Keith Waldron[3] and Ian Bancroft[1,4*]

Does the title give you information?
What does it tell you?

Letter2: Miller 2016

abstract

Note how the abstract has subsections.

Can you identify these sections in the abstract:

1. **Hypothesis or question**
2. **Experiment or test**
3. **Result(s)**
4. **Interpretations/conclusions**

Abstract

Background: The current approach to reducing the tendency for wheat grown under high fertilizer conditions to collapse (lodge) under the weight of its grain is based on reducing stem height via the introduction of *Rht* genes. However, these reduce the yield of straw (itself an important commodity) and introduce other undesirable characteristics. Identification of alternative height-control loci is therefore of key interest. In addition, the improvement of stem mechanical strength provides a further way through which lodging can be reduced.

Results: To investigate the prospects for genetic alternatives to *Rht*, we assessed variation for plant height and stem strength properties in a training genetic diversity panel of 100 wheat accessions fixed for *Rht*. Using mRNAseq data derived from RNA purified from leaves, functional genotypes were developed for the panel comprising 42,066 Single Nucleotide Polymorphism (SNP) markers and 94,060 Gene Expression Markers (GEMs). In the first application in wheat of the recently-developed method of Associative Transcriptomics, we identified associations between trait variation and both SNPs and GEMs. Analysis of marker-trait associations revealed candidates for the causative genes underlying the trait variation, implicating xylan acetylation and the COP9 signalosome as contributing to stem strength and auxin in the control of the observed variation for plant height. Predictive capabilities of key markers for stem strength were validated using a test genetic diversity panel of 30 further wheat accessions.

Conclusions: This work illustrates the power of Associative Transcriptomics for the exploration of complex traits of high agronomic importance in wheat. The careful selection of genotypes included in the analysis, allowed for high resolution mapping of novel trait-controlling loci in this staple crop. The use of Gene Expression markers coupled with the more traditional sequence-based markers, provides the power required to understand the biological context of the marker-trait associations observed. This not only adds to the wealth of knowledge that we strive to accumulate regarding gene function and plant adaptation, but also provides breeders with the information required to make more informed decisions regarding the potential consequences of incorporating the use of particular markers into future breeding programmes.

Keywords: Modulus of Rupture, lodging, Associative Transcriptomics, Xylan acetylation, COP9 signalosome, Auxin

Letter2: Miller 2016

abstract

Note how the abstract has subsections.

Can you identify these sections in the abstract:

1. **Hypothesis or question**
2. **Experiment or test**
3. **Result(s)**
4. **Interpretations/conclusions**

Abstract

1

Background: The current approach to reducing the tendency for wheat grown under high fertilizer conditions to collapse (lodge) under the weight of its grain is based on reducing stem height via the introduction of *Rht* genes. However, these reduce the yield of straw (itself an important commodity) and introduce other undesirable characteristics. Identification of alternative height-control loci is therefore of key interest. In addition, the improvement of stem mechanical strength provides a further way through which lodging can be reduced.

2

Results: To investigate the prospects for genetic alternatives to *Rht*, we assessed variation for plant height and stem strength properties in a training genetic diversity panel of 100 wheat accessions fixed for *Rht*. Using mRNAseq data derived from RNA purified from leaves, functional genotypes were developed for the panel comprising 42,066 Single Nucleotide Polymorphism (SNP) markers and 94,060 Gene Expression Markers (GEMs). In the first application in wheat of the recently-developed method of Associative Transcriptomics, we identified associations between trait variation and both SNPs and GEMs. Analysis of marker-trait associations revealed candidates for the causative genes underlying the trait variation, implicating xylan acetylation and the COP9 signalosome as contributing to stem strength and auxin in the control of the observed variation for plant height. Predictive capabilities of key markers for stem strength were validated using a test genetic diversity panel of 30 further wheat accessions.

3

4

Conclusions: This work illustrates the power of Associative Transcriptomics for the exploration of complex traits of high agronomic importance in wheat. The careful selection of genotypes included in the analysis, allowed for high resolution mapping of novel trait-controlling loci in this staple crop. The use of Gene Expression markers coupled with the more traditional sequence-based markers, provides the power required to understand the biological context of the marker-trait associations observed. This not only adds to the wealth of knowledge that we strive to accumulate regarding gene function and plant adaptation, but also provides breeders with the information required to make more informed decisions regarding the potential consequences of incorporating the use of particular markers into future breeding programmes.

Keywords: Modulus of Rupture, lodging, Associative Transcriptomics, Xylan acetylation, COP9 signalosome, Auxin

Letter2: Miller 2016

How to describe results

Part 1: this section describes.

Results

Variation for stem structural and material strength

The diversity panel of 100 wheat accessions was analysed for a range of traits indicative of stem structural and material strength. With the exception of second moment of area, significant variation was present for all traits included in the analysis ($P < 0.05$) (Additional file 1). The absolute strength traits Fmax and F/V showed respective trait ranges of 7.45–38.55 and 29.82–80.44 N/s. The wheat accession displaying highest stem absolute strength (for both Fmax and F/V) was Orlando. The lowest trait values were seen in Battalion and Escorial for F/V and Fmax respectively. For the material strength traits, MOR and MOE, respective trait ranges of 0.70–8.05 and 121.6–1490.3 Nmm^{-2} were recorded. Of the wheat accessions screened, Gatsby exhibited the lowest trait values for both MOE and MOR. Accessions displaying the highest material strength were Alba (for MOR) and Cordiale (for MOE). A wide range of variation was also observed for the various stem structural traits assessed. For example, mean stem hollow area ranged from 1.16 mm^2 (for Capelle-Desprez) and 6.51 mm^2 (for Starke2). For outer cortex thickness, trait means ranging between 0.24 mm (as seen for Hyperion) and 0.46 mm (as seen for Alba) were recorded. For plant height, despite a lack of segregation at the *Rht* loci, a trait range of 42.8–98.4 cm was recorded. The tallest accession included within the panel was Steadfast whereas the shortest stem measurements were recorded for Equinox.

Letter2: Miller 2016

How to describe results

A correlation analysis was performed to analyse the relationships between the absolute strength and the structural and morphological traits to assess which may be good breeding targets (Table 1). Several highly significant ($P \leq 0.001$) relationships were detected between the absolute strength measures (Fmax and F/V) and the structural traits, however, despite such high statistical significance, in the majority of cases, the amount of variation in stem absolute strength explained by stem structure was found to be modest. Stem parenchyma area ($R^2 = 0.27$ and 0.17 for Fmax and F/V respectively) and outer cortex thickness ($R^2 = 0.19$ and 0.13 for Fmax and F/V respectively) show the closest positive relationships with absolute strength. These traits may therefore be the most promising targets for the improvement of stem structural strength in wheat. In contrast to the modest contributions made by stem geometry, a much closer correlation is seen between the absolute strength measures and stem weight ($R^2 = 0.42$ and 0.47 for Fmax and F/V respectively). These correlations may represent a combined effect of several different stem structural components (each contributing to weight) or may more specifically relate to the density of the materials that make up the plant stem. Plant height also correlates positively with stem absolute strength ($R^2 = 0.21$ and 0.25 for Fmax and F/V respectively).

The lack of strong correlations observed between stem structure and absolute strength may suggest that stem material properties are of high value for the improvement of stem mechanical strength in wheat. Consistent with this, the relationship between the field-based measure of stem lodging risk (utilising the pulley system illustrated in Fig. 1c) and the absolute and material strength traits, showed a stronger correlation for the material strength trait Modulus of Rupture (MOR; R^2 of 0.41, $P < 0.001$) in comparison to absolute strength traits such as Fmax (R^2 of 0.27, $P < 0.001$) (Additional file 4).

Part 2: what does this section do?
Why did we need part 1?

Letter2: Miller 2016

How to describe results

Associative Transcriptomics for plant height

In order to identify loci controlling plant height, AT was conducted using the functional genotypes scored and the plant height trait data obtained. Additional file 6 summarises the results obtained. Two major association peaks were identified: one on chromosome 6A and the other on 5B, each exhibiting SNP and GEM associations (Fig. 2). To identify candidates for the causative genes for control of the trait underlying the association peaks, the sequence similarities of unigenes to gene models in Brachypodium, Sorghum, rice and Arabidopsis were used as a guide to gene function. This revealed that the gene corresponding to the highest significance GEM on 6A is an orthologue of a rice Auxin Response Factor (*OsARF16*, Os02g41800; Panel a). The peak found on chromosome 5B coincided with a cluster of *SMALL AUXIN UP RNA* (SAUR) genes, with high significance GEMs occurring in three of the unigenes with BLAST identity to SAUR genes (Panel b). Although these loci have not been implicated previously in the control of plant height in wheat, the genes identified are excellent candidates for controlling this trait: ARFs are transcription factors that bind specifically to auxin response elements (*AuxREs*) found in the promoters of early auxin response genes such as the large family of SAUR genes, and mediate their response to auxin [20]. In wheat, we found that the GEM for the ARF on 6A had a positive correlation with stem height. These results suggest that this Auxin Response Factor may have a developmental role in wheat. Although the actual function of the SAURs is not known, it has been reported that some have an important role in control of cell expansion and patterning [21]. On closer inspection of their sequence similarities, the SAUR genes in the region of 5B are putative orthologues of some of the members of a cluster of 17 SAURs found on rice chromosome 9 (*OsSAUR39-55*) and an orthologous cluster can also be found on Arabidopsis chromosome 1 (*AtSAUR61-68*) [22]. In rice, *OsSAUR39* has been found to negatively regulate auxin synthesis and transport, leading to reduced growth phenotypes when over-expressed [23]. Our observation that all of the highly associated SAURs in this cluster exhibited gene expression that was negatively correlated with height is concordant with this.

Identify these parts:

1. **Hypothesis or question**
2. **Experiment or test**
3. **Results**
4. **Interpretations, if any**

NB: This section runs through this pattern **twice** (sort of).

Letter2: Miller 2016

How to describe results

Associative Transcriptomics for plant height

1 In order to identify loci controlling plant height, AT was conducted using the functional genotypes scored and the plant height trait data obtained. Additional file 6 summarises the results obtained. Two major association peaks were identified: one on chromosome 6A and the other on 5B, each exhibiting SNP and GEM associations (Fig. 2). To identify candidates for the causative genes for control of the trait underlying the association peaks, 2 the sequence similarities of unigenes to gene models in Brachypodium, Sorghum, rice and Arabidopsis were used as a guide to gene function. This revealed that the gene corresponding to the highest significance GEM on 6A is an orthologue of a rice Auxin Response Factor (*OsARF16*, Os02g41800; Panel a). The peak found on chromosome 5B coincided with a cluster of *SMALL AUXIN UP RNA* (SAUR) genes, with high significance GEMs occurring in three of the unigenes with BLAST identity to SAUR genes (Panel b). Although these loci have not been implicated previously in the control of plant height in wheat, the genes identified are excellent candidates for controlling this trait: ARFs are transcription factors that bind specifically to auxin response elements (*AuxREs*) found in the promoters of early auxin response genes such as the large family of SAUR genes, and mediate their response to auxin [20]. In wheat, we found that the GEM for the ARF on 6A had a positive correlation with stem height. These results suggest that this Auxin Response Factor may have a developmental role in wheat. Although the actual function of the SAURs is not known, it has been reported that some have an important role in control of cell expansion and patterning [21]. On closer inspection of their sequence similarities, the SAUR genes in the region of 5B are putative orthologues of some of the members of a cluster of 17 SAURs found on rice chromosome 9 (*OsSAUR39-55*) and an orthologous cluster can also be found on Arabidopsis chromosome 1 (*AtSAUR61-68*) [22]. In rice, *OsSAUR39* has been found to negatively regulate auxin synthesis and transport, leading to reduced growth phenotypes when over-expressed [23]. Our observation that all of the highly associated SAURs in this cluster exhibited gene expression that was negatively correlated with height is concordant with this.

3

Identify these parts:

1. **Hypothesis or question**
2. **Experiment or test**
3. **Results**
4. **Interpretations, if any**

Letter2: Miller 2016

Part 4: The figure.

Why use a figure for this result?

Why use arrows in the figure?

What is good or bad about the figure legend?

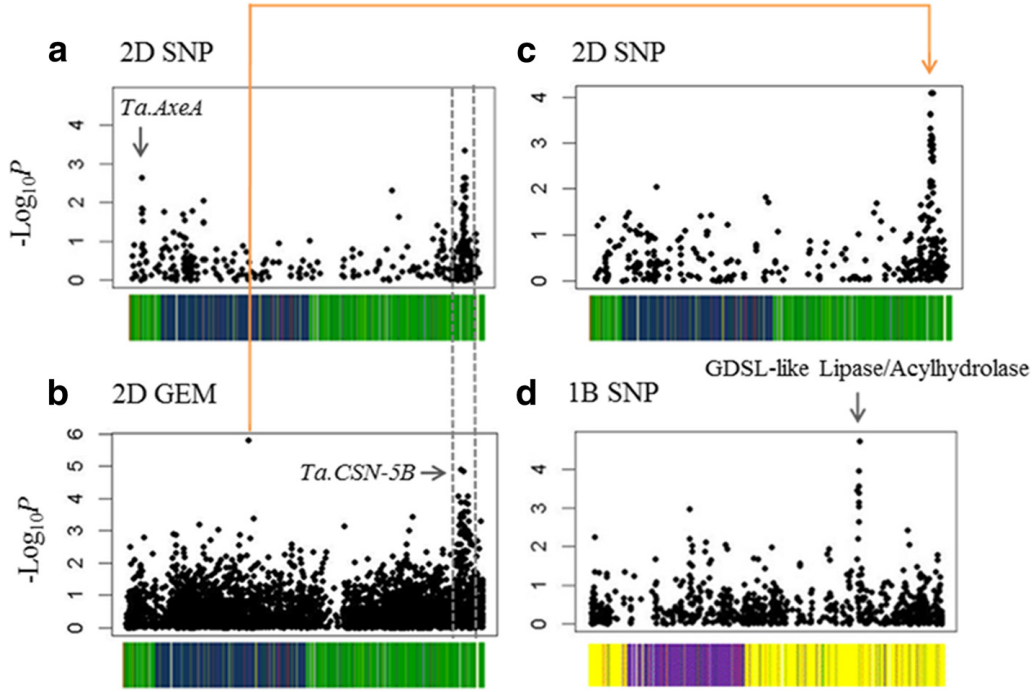

Fig. 3 Variation at both the sequence (SNP) and gene expression (GEM) level show high association with MOR. Two SNP association peaks for MOR were seen on chromosome 2D (**a**). The peak to the right of panel a was also identified in the GEM analysis (**b**). Several single GEM associations were also detected for MOR (see single GEM at the foot of the orange line in panel **b** as an example). Mapping transcript abundanc (as RPKM) as a trait against the SNP data revealed the same 2D SNP peak for several single GEMs (see panel **c** for an example). A further SNP association for MOR was detected on chromosome 1B **(d)**. The positions of candidate genes are indicated by *arrows*. -Log10*P* values are plotted in wheat pseudomolecule order. Unigene order is colour-coded according to sequence similarity to *B. distachyon* chromosomes (*blue* = Bd1; *yellow* = Bd2; *purple* = Bd3; *red* = Bd4 and *green* = Bd5). Position of candidate genes are indicated by arrows

Explaining results: the exercise

Practice at explaining results, using the yeast data:

Hypothesis or question:

Are the two gene expression measures consistent?

Test/analysis: discuss how to explain this

Result: discuss how to explain this

Interpretation: discuss how to explain this

```
#load the yeast data
load("fission_yeast_data.10-Feb-2020.Rda")
#or load from the previous sessions

#make a plot, using two log scales
plot(
  log10(gene$mRNA_copies_per_cell),
  log10(gene$gene.expression.RPKM)
)
#check the correlation
cor.test(
  gene$mRNA_copies_per_cell,
  gene$gene.expression.RPKM,
 method="spearman"
)
```

Summary: the question(s)

The big question: what are you trying to find out?

What organism?
What process?
What data will you use?
How will you analyse that data?

Think about this before you start analysing data!

Summary: introduction = background

The introduction sets the scene:

- Gives the broad context (eg: *Brassica napus* is, glucosinolates are).
- Explains what is known already (with references).
- Explains what you want to find out.
- May explain the method, very briefly.

Summary: results are important!

The results are the MOST IMPORTANT PART.

Each results and discussion section should contain:

1. **A hypothesis or question**
2. **An experiment or a test**
3. **A description of results: a plot can help!**
4. **Interpretations, if any.**

Summary: the conclusion

The conclusion should briefly reiterate the major findings of your study.

- **It should tell us what you found.**
- **It should be short (don't ramble).**
- **It should mention caveats or limitations.**

Summary: advice

Welcome to Big Data Biology (BIO00047I)

Quick links: Module synopsis | VLE Site | List of staff | Description of the data | Report Guide

1. Read these

The assessement

The assessment for this module is a report describing an analysis of data (maximum 1500 words). A description of what we want to see, and how we will mark the report is here

The assessment deadline is: Thursday 16th of April 2020 at 11am

2. Talk to us about your project plan.

3. Do NOT leave it to the last minute!
Finish it three days before.
Have a rest.
Re-read and correct your grammar and spelling errors.