



# Article Underestimation of Heritability across the Molecular Layers of the Gene Expression Process

Jihye Ryu and Chaeyoung Lee \*D

School of Systems Biomedical Science, Integrative Institute of Basic Science, Soongsil University, 369 Sangdo-ro, Dongjak-gu, Seoul 06978, Korea; toriyah@ssu.ac.kr

\* Correspondence: clee@ssu.ac.kr; Tel.: +82-2-820-0455

Abstract: We investigated the extent of the heritability underestimation for molecules from an infinitesimal model in mixed model analysis. To this end, we estimated the heritability of transcription, ribosome occupancy, and translation in lymphoblastoid cell lines from Yoruba individuals. Upon considering all genome-wide nucleotide variants, a considerable underestimation in heritability was observed for mRNA transcription (-0.52), ribosome occupancy (-0.48), and protein abundance (-0.47). We employed a mixed model with an optimal number of nucleotide variants, which maximized heritability, and identified two novel expression quantitative trait loci (eQTLs;  $p < 1.0 \times 10^{-5}$ ): rs11016815 on chromosome 10 that influences the transcription of SCP2, a trans-eGene on chromosome 1-whose expression increases in response to MGMT downregulation-induced apoptosis, the cis-eGene of rs11016815—and rs1041872 on chromosome 11 that influences the ribosome occupancy of CCDC25 on chromosome 8 and whose cis-eGene encodes ZNF215, a transcription factor that potentially regulates the translation speed of CCDC25. Our results suggest that an optimal number of nucleotide variants should be used in a mixed model analysis to accurately estimate heritability and identify eQTLs. Moreover, a heterogeneous covariance structure based on gene identity and the molecular layers of the gene expression process should be constructed to better explain polygenic effects and reduce errors in identifying eQTLs.

**Keywords:** expression quantitative trait locus; gene regulation; heritability; *MGMT*; mixed model; protein abundance; ribosome occupancy; *ZNF215* 

# 1. Introduction

Gene expression is a critical process that links genetic information to phenotypes. Nucleotide sequence variants that are associated with gene expression are called expression quantitative trait loci (eQTLs), and genome-wide eQTL analyses help to dissect the genetic mechanism underlying gene regulation by generating *cis*- and *trans*-eQTL profiles for a gene [1]. A substantial proportion of gene expression variation was attributed to eQTLs [2,3]. The eQTLs were heterogeneous by ethnic groups [2], which indicates potential spurious genetic associations produced by the population structure [4]. Moreover, the polygenic effects of gene expression have been largely observed [3], and a great concern for the accuracy of eQTLs was raised by ignoring polygenic effects in the eQTL analyses [5]. Both the population structure and polygenic effects can be explained in the genome-wide analysis by a mixed model, which reduces false positive and negative eQTLs [6]. The mixed model uses a genomic covariance structure among individuals to explain random polygenic effects; thus, the polygenic effects in the analytical model reflect all the variants across the genome simultaneously [5]. For a mixed model analysis with pedigree information, an infinitesimal model is conventionally assumed [7,8]; however, the assumption that gene expression is regulated by all genetic variants is unrealistic, especially when applied to every gene. This concern was raised in a simulation study, wherein the infinitesimal model led to an underestimation of heritability [9]. This was a consequence of the inclusion



Citation: Ryu, J.; Lee, C. Underestimation of Heritability across the Molecular Layers of the Gene Expression Process. *Processes* 2021, *9*, 2144. https://doi.org/ 10.3390/pr9122144

Academic Editor: Bonglee Kim

Received: 16 November 2021 Accepted: 25 November 2021 Published: 27 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). of an excessive number of non-effective variants, which include the regulatory sequences for genes not expressed in a specific cell. Moreover, non-effective variants can be heterogeneous with respect to the molecular layers of gene expression, namely transcription, ribosome occupancy, and translation.

In this study, we aimed to examine the extent of the underestimation of heritability of gene expression under the infinitesimal model using a mixed model and to assess the heterogeneity of heritability across the different molecular layers of the gene expression process. This study illustrates, for the first time, false-negative eQTLs with an underestimated heritability of gene expression by employing an infinitesimal model. Our results indicate the need for the construction of an optimal genomic covariance structure for genome-wide eQTL analyses.

#### 2. Materials and Methods

### 2.1. Expression and Genotype Data

To estimate the heritability of mRNA transcription, ribosome occupancy, and protein abundance, we used gene expression data from lymphoblastoid cell lines derived from unrelated Yoruba individuals from Ibadan, Nigeria [10,11]. The expression levels of the mRNA transcript and the ribosome occupancy were quantified as a logarithm of read counts per kilobase per million mapped reads of each gene for each individual using the Illumina Genome Analyzer 2 platform [10] and an Illumina HiSeq 2500 [11], respectively. The protein abundance level was quantified as a logarithm of ratio to the stable isotope labeling by amino acids in cell culture (SILAC) internal standard sample of each gene for each individual using protein mass spectrometry [11]. The expression data were standardized and quantile-normalized for each molecular layer of each gene to fit a standard normal distribution. This study used the expression data of 2246 genes that are commonly available for mRNA transcription, ribosome occupancy, and protein abundance.

The genotype data of the Yoruba individuals were obtained from the 1000 Genomes Project Phase 3 (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/, accessed on 15 May 2018). After filtering out nucleotide sequence variants using the Hardy–Weinberg disequilibrium criteria ( $p < 1.0 \times 10^{-6}$ ) and a minor allele frequency <0.1, the genotype data of 5,594,467 variants remained. Finally, the heritability was estimated using 63, 62, and 51 individuals for mRNA transcription, ribosome occupancy, and protein abundance, respectively.

### 2.2. Statistical Analysis

We employed a mixed model to estimate heritability as follows:

$$y = g + \varepsilon$$

where y is the vector for gene expression levels for mRNA transcription, ribosome occupancy, or protein abundance, g is the vector for random polygenic effects with  $g \sim N(0, A\sigma_g^2)$ , and  $\varepsilon$  is the vector for random residuals with  $\varepsilon \sim N(0, I\sigma_{\varepsilon}^2)$ .  $\sigma_g^2$  is the polygenic variance component, and  $\sigma_{\varepsilon}^2$  is the environmental variance component. A is the genomic similarity matrix (GSM), and I is the identity matrix. The GSM has elements of pairwise genomic similarity coefficients estimated using sequence variants, all or a part of them with a significance threshold ranging from  $1.0 \times 10^{-6}$  to  $5.0 \times 10^{-2}$ . The genomic similarity coefficient between individuals j and k was calculated as follows:

$$g_{jk} = \frac{1}{n_v} \sum_{i=1}^{n_v} \frac{(\tau_{ij} - 2f_i)(\tau_{ik} - 2f_i)}{2f_i(1 - f_i)}$$

where  $n_v$  is the number of nucleotide variants that contribute to the genomic similarity,  $\tau_{ij}$  and  $\tau_{ik}$  represent the number (0, 1, or 2) of minor alleles for the nucleotide variant *i*, and  $f_i$  is the frequency of the minor allele.

The variance components of  $\sigma_g^2$  and  $\sigma_{\varepsilon}^2$  were estimated using the AI-REML algorithm, and the heritability was calculated as  $\frac{\sigma_g^2}{\sigma_g^2 + \sigma_{\varepsilon}^2}$ . All estimates were obtained using Genomewide Complex Trait Analysis (GCTA, v1.26; http://cnsgenomics.com/software/gcta/, accessed on 10 July 2018).

An optimal GSM for mRNA transcription, ribosome occupancy, and protein abundance was proposed to maximize heritability. A mixed model analysis with the optimal GSM was conducted to identify novel eQTLs that were not identified with the conventional GSM that considered all the nucleotide variants. We identified eQTLs using the following analytical model including random polygenic effects with GSM:

$$y = \mu \mathbf{1} + x\beta + g + \epsilon$$

where  $\mu$  is the overall mean, **1** is the vector of 1's,  $\beta$  is the fixed minor allele effect of the nucleotide variant to be tested for association, and x is the vector with elements of 0, 1, and 2 for the homozygote of the major allele, the heterozygote, and the homozygote of the minor allele, respectively. The vectors of random variables (y, g, and  $\varepsilon$ ) are the same as above. The eQTL analysis was conducted with optimal and conventional GSMs, and eQTLs. Thus, the two different definitions of g are as follows:  $g^o \sim N(0, A^o \sigma_g^2)$  and  $g^c \sim N(0, A^c \sigma_g^2)$ , where  $A^o$  is the optimal GSM, and  $A^c$  is the conventional GSM. When the optimal GSM was calculated, we included the nucleotide variants with eQTLs that were proposed to maximize the heritability of each molecular layer of the expression process. On the other hand, the conventional GSM considered all the nucleotide variants. However, on the construction of the GSM for both analyses, we excluded all the nucleotide variants on a chromosome where the candidate eQTL to be tested was located. This helped avoid the underestimation of the eQTL effect and reduce false negatives. The variance components were estimated using the AI-REML algorithm, and the allele effect was then estimated and tested.

## 3. Results

The heritability of gene expression was estimated by considering eQTLs with various significance thresholds for mRNA transcription, ribosome occupancy, and protein abundance, and their average values are presented in Figure 1. The maximum estimate of heritability was approximately 1 (0.99) for each molecular layer examined. The maximum value resulted from the analysis wherein GSM was calculated by considering eQTLs with  $p < 6.0 \times 10^{-4}$ ,  $p < 1.0 \times 10^{-3}$ , and  $p < 1.0 \times 10^{-3}$  for mRNA transcription, ribosome occupancy, and protein abundance, respectively. Additionally, including all the variants in the analysis resulted in smaller values of heritability of 0.47, 0.51, and 0.52 for mRNA transcription, ribosome occupancy, and protein abundance, respectively (Figure 1).

We compared the heritability of mRNA transcription, ribosome occupancy, and protein abundance and found that the heritability of protein abundance was higher than that for mRNA transcription when the significance threshold was  $1.0 \times 10^{-5}$  or less. On the contrary, this difference was smaller when a larger significance threshold ( $2.0 \times 10^{-5}$ to  $1.0 \times 10^{-3}$ ) was used. The minimum difference between the heritability of mRNA transcription and that of protein abundance was observed at a significance threshold of  $5.0 \times 10^{-6}$ , while the maximum difference was noted at a significance threshold of  $9.6 \times 10^{-5}$  (Table 1). The number of protein abundance eQTLs (pQTLs) was larger than and lower than that of narrow-sense eQTLs (neQTLs), at the minimum and the maximum differences, respectively. However, the average proportion explained by one pQTL was larger than that explained by one neQTL, regardless of the significance threshold values.



**Figure 1.** The heritability of mRNA transcription, ribosome occupancy, and protein abundance according to different significance thresholds of quantitative trait loci.  $\Delta_{ne}$ ,  $\Delta_{r}$ , and  $\Delta_{p}$  represent the differences between the maximum heritability and the heritability considering all the variants across the genome for mRNA transcription, ribosome occupancy, and protein abundance, respectively.

**Table 1.** The characteristics at the minimum and maximum differences in heritability between mRNA transcription and protein abundance, explained by expression quantitative trait loci (eQTLs).

	Features	mRNA Transcription (A)	Protein Abundance (B)	Difference (A-B)	<i>p</i> -Value
d <sub>min</sub> *	$h^2$ No. of eQTLs $h^2$ /No. of eQTLs	0.477 13.9 0.221	0.525 16.9 0.281	$-0.048 \\ -3.0 \\ -0.060$	$egin{array}{llllllllllllllllllllllllllllllllllll$
d <sub>max</sub> *	$h^2$ No. of eQTLs $h^2$ /No. of eQTLs	0.801 138.2 0.011	0.699 75.6 0.044	$0.102 \\ 62.6 \\ -0.033$	$\begin{array}{c} 1.32\times 10^{-156}\\ 9.56\times 10^{-46}\\ 4.80\times 10^{-151}\end{array}$

\*  $d_{min}$  and  $d_{max}$  indicate the minimum and maximum h<sup>2</sup> differences between mRNA transcription and protein abundance.  $d_{min}(d_{max})$  were observed when eQTLs with  $p < 5.0 \times 10^{-6}$  ( $p < 9.6 \times 10^{-5}$ ) were considered for constructing the genomic similarity matrix. h<sup>2</sup>, heritability; eQTL, expression quantitative trait locus.

When we employed the GSM that resulted in the maximum heritability of gene expression, two novel eQTLs, rs11016815 (neQTL) and rs1041872 (ribosome occupancy eQTL; rQTL), were identified for *SCP2* and *CCDC25* (Table 2,  $p < 1.0 \times 10^{-5}$ ). These are both *trans*-acting eQTLs.

**Table 2.** Novel expression quantitative trait loci (eQTLs) resulted from a mixed model analysis employing the proposed genome similarity matrix \*.

eQTL Type	SNP	Position §	Allele <sup>¶</sup>	MAF	eGene (Chr)	cis/trans Regulation	Significance w/ ‡	
							Conventional GSM	Optimal GSM
neQTL	rs11016815	10; 129502351	G/C	0.1032	SCP2 (1)	trans	$2.84 imes10^{-3}$	$2.98  imes 10^{-6}$
rQTL	rs1041872	11; 6979775	G/A	0.1048	CCDC25 (8)	trans	$3.40  imes 10^{-3}$	$9.46\times10^{-6}$

\* The nucleotide sequence variants with  $p < 1.0 \times 10^{-5}$  are presented. <sup>‡</sup> The conventional GSM was estimated using 5,594,467 nucleotide variants across the genome. The optimal GSM was estimated considering eQTLs with  $p < 6.0 \times 10^{-4}$  and  $p < 1.0 \times 10^{-4}$  for mRNA transcription and ribosome occupancy, respectively. <sup>§</sup> The chromosomal position (Chr; bp) was obtained from the GRCh38 reference genome. <sup>¶</sup> Major/minor allele. GSM, genomic similarity matrix; SNP, single nucleotide polymorphism; MAF, minor allele frequency; eQTL, expression quantitative trait locus; rQTL, ribosome occupancy eQTL; neQTL, narrow-sense eQTL.

#### 4. Discussion

This study showed the gene expression heritability explained by eQTLs, considered for the polygenic covariance structure among individuals in a mixed model framework.

As expected, we found that the heritability estimates increased as the number of significant eQTLs increased. The heritability estimates were, however, underestimated when all the genome-wide nucleotide variants were included in the analysis. Polygenic effects might be diluted with the undesirable covariance structure reflecting non-effective eQTLs. As a result, the heritability was underestimated. The heritability was maximum when eQTLs with  $p < 6.0 \times 10^{-4}$ ,  $p < 1.0 \times 10^{-3}$ , and  $p < 1.0 \times 10^{-3}$  were considered for mRNA transcription, ribosome occupancy, and protein abundance, respectively. The maximum estimates were almost equivalent to one (>0.99), indicating that the considered eQTLs almost fully explained the variability in gene expression. A considerable reduction in heritability ( $\Delta_{ne} = -0.52$ ,  $\Delta_{r} = -0.48$ , and  $\Delta_{p} = -0.47$ ) was found by including the nucleotide variants across the whole genome. This underestimation concurred with a previous study in which such an underestimation was also found in the heritability explained by QTLs for phenotypes [9]. However, the enormous underestimation for the gene expression may not be negligible in contrast to the underestimation for phenotypes (e.g.,  $\Delta = -0.02$  for hypertension). The assumption that gene expression variability is a consequence of all genetic variants should be avoided. Our results suggest that an optimal GSM should be employed for the mixed model analysis of eQTLs.

Moreover, an optimal GSM customized for each molecular layer of the expression process should be used to account for heterogeneous polygenic effects, even though the same gene is being analyzed across the different molecular layers. As shown above, a larger underestimation was found for mRNA transcription compared to that for ribosome occupancy and protein abundance. Correspondingly, the number of eQTLs considered in constructing the GSM for mRNA transcription (1242 neQTLs) was smaller than that for ribosome occupancy (1604 rQTLs) and protein abundance (1540 pQTLs). The use of a GSM heterogenous across molecular layers is critical for genes that are heavily controlled at some specific step(s) in gene expression. For example, the gene expression of ribosomal proteins is mainly controlled by pQTLs during translation in response to the instantaneous changes in the cell environment [12].

The current study revealed two novel eQTLs using the optimal GSM in mixed model analysis. This indicates that false negatives were produced as a result of a diluted genomic background, wherein polygenic effects were confounded by the presence of non-effective variants in an infinitesimal model. This indicates that the underestimated polygenic variance components lead to the identification of inaccurate individual eQTL effects across the genome.

The novel eQTLs were identified as *trans*-acting neQTLs and rQTLs of the SCP2 and CCDC25 genes. They are both intragenic variants (rs11016815 and rs1041872) in the MGMT and ZNF215 genes, respectively. The MGMT might be a cis-eGene of the rs11016815, which was supported by previous studies in which this variant has been found to be associated with the transcription levels of *MGMT* in the blood ( $p = 5.4 \times 10^{-11}$ ; [13]) and pancreas ( $p = 2.8 \times 10^{-7}$ ; GTEx Consortium, 2020) [14]. A potential influence of *MGMT* on the SCP2 gene expression is suspected. This is because MGMT can repair damaged guanine nucleotides and prevent apoptosis [15], and the SCP2 gene increases lipid transfer during apoptosis [16]. The evidence that ZNF215 might be a *cis*-eGene of rs1041872 arises from their association ( $p = 7.2 \times 10^{-6}$ ) in the GTEx Consortium [14]. Because the functions of ZNF215 are relatively unknown [17], it is difficult to infer the functional influence on the CCDC25 gene [18]. However, it is highly likely that ZNF215 functions as a transcription factor because this zinc finger protein has motifs that commonly exist in transcription factors. Its C2H2 zinc finger motif can interact directly with DNA, and this is the most common characteristic of transcription factors in humans [19]. Its KRAB region can bind to the KAP1 co-repressor as a transcription repressor domain [20]. Moreover, a study observed that ZNF215 was transported to the nucleus using GFP fusion proteins [21]. We suspect that ZNF215 acts as a *cis*-eGene of rs1041872 and might be involved in trans-regulating the speed of the translational elongation of the CCDC25 gene.

# 5. Conclusions

This study demonstrated the inaccurate identification of eQTLs produced by employing a GSM under the infinitesimal model and that an optimal GSM was critical for accurate eQTL mapping. We suggest that an optimal GSM should be constructed based on the different molecular layers of gene expression as well as the identity of the gene. More accurate eQTL profiles would help understand the genetics of complex diseases and ultimately contribute to precision medicine.

**Author Contributions:** Conceptualization, C.L.; formal analysis, J.R.; writing, C.L. and J.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Ministry of Education (2021R1A6A1A10044154).

**Institutional Review Board Statement:** This study was exempt from IRB review because we used publicly available population data and subjects could not be identified.

Informed Consent Statement: Not applicable.

**Data Availability Statement:** The data used in this study are publicly available in the GEO DataSets (Accession No. GSE61742).

**Conflicts of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

#### References

- Brynedal, B.; Choi, J.; Raj, T.; Bjornson, R.; Stranger, B.E.; Neale, B.M.; Voight, B.F.; Cotsapas, C. Large-scale *trans*-eQTLs affect hundreds of transcripts and mediate patterns of transcriptional co-regulation. *Am. J. Hum. Genet.* 2017, 100, 581–591. [CrossRef] [PubMed]
- Spielman, R.S.; Bastone, L.A.; Burdick, J.T.; Morley, M.; Ewens, W.J.; Cheung, V.G. Common genetic variants account for differences in gene expression among ethnic groups. *Nat. Genet.* 2007, *39*, 226–231. [CrossRef] [PubMed]
- 3. Yang, S.; Liu, Y.; Jiang, N.; Chen, J.; Leach, L.; Luo, Z.; Wang, M. Genome-wide eQTLs and heritability for gene expression traits in unrelated individuals. *BMC Genom.* 2014, *15*, 1–12. [CrossRef]
- 4. Gay, N.R.; Gloudemans, M.; Antonio, M.L.; Abell, N.S.; Balliu, B.; Park, Y.; Martin, A.R.; Musharoff, S.; Rao, A.S.; Aguet, F.; et al. Impact of admixture and ancestry on eQTL analysis and GWAS colocalization in GTEx. *Genome Biol.* **2020**, *21*, 1–20. [CrossRef]
- 5. Lee, C. Genome-wide expression quantitative trait loci analysis using mixed models. Front. Genet. 2018, 9, 341. [CrossRef]
- 6. Shin, J.; Lee, C. A mixed model reduces spurious genetic associations produced by population stratification in genome-wide association studies. *Genomics* **2015**, *105*, 191–196. [CrossRef] [PubMed]
- Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: A tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 2011, 88, 76–82. [CrossRef]
- Carayol, J.; Chabert, C.; Di Cara, A.; Armenise, C.; Lefebvre, G.; Langin, D.; Viguerie, N.; Metairon, S.; Saris, W.H.M.; Astrup, A.; et al. Protein quantitative trait locus study in obesity during weight-loss identifies a leptin regulator. *Nat. Commun.* 2017, *8*, 2084. [CrossRef] [PubMed]
- 9. Ryoo, H.; Lee, C. Underestimation of heritability using a mixed model with a polygenic covariance structure in a genome-wide association study for complex traits. *Eur. J. Hum. Genet.* **2014**, *22*, 851–854. [CrossRef] [PubMed]
- 10. Pickrell, J.K.; Pai, A.A.; Gilad, Y.; Pritchard, J.K. Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet.* **2010**, *6*, e1001236. [CrossRef] [PubMed]
- 11. Battle, A.; Khan, Z.; Wang, S.H.; Mitrano, A.; Ford, M.J.; Pritchard, J.K.; Gilad, Y. Genomic variation. Impact of regulatory variation from RNA to protein. *Science* **2015**, *347*, 664–667. [CrossRef]
- 12. Ryu, J.; Lee, C. Regulatory nucleotide sequence signals for expression of the genes encoding ribosomal proteins. *Front. Genet.* **2020**, *11*, 501. [CrossRef]
- Jansen, R.; Hottenga, J.J.; Nivard, M.G.; Abdellaoui, A.; Laport, B.; de Geus, E.J.; Wright, F.A.; Penninx, B.W.J.H.; Boomsma, D.I. Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Hum. Mol. Genet.* 2017, 26, 1444–1451. [CrossRef] [PubMed]
- 14. GTEx Consortium. The GTEx consortium atlas of genetic regulatory effects across human tissues. GT Sci. 2020, 369, 1318–1330.
- 15. Kaina, B.; Christmann, M.; Naumann, S.; Roos, W.P. MGMT: Key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. *DNA Repair* **2007**, *6*, 1079–1099. [CrossRef] [PubMed]
- 16. Liang, R.; Shen, X.L.; Zhang, B.; Li, Y.; Xu, W.; Zhao, C.; Luo, Y.; Huang, K. Apoptosis signal-regulating kinase 1 promotes ochratoxin A-induced renal cytotoxicity. *Sci. Rep.* **2015**, *5*, 8078. [CrossRef]
- 17. Massart, F.; Saggese, G. Morphogenetic targets and genetics of undescended testis. Sex Dev. 2010, 4, 326–335. [CrossRef]

- 18. Yang, L.; Liu, Q.; Zhang, X.; Liu, X.; Zhou, B.; Chen, J.; Huang, D.; Li, J.; Li, H.; Chen, F.; et al. DNA of neutrophil extracellular traps promotes cancer metastasis via CCDC25. *Nature* **2020**, *583*, 133–138. [CrossRef]
- 19. Vaquerizas, J.M.; Kummerfeld, S.K.; Teichmann, S.A.; Luscombe, N.M. A census of human transcription factors: Function, expression and evolution. *Nat. Rev. Genet.* **2009**, *10*, 252–263. [CrossRef] [PubMed]
- Lupo, A.; Cesaro, E.; Montano, G.; Zurlo, D.; Izzo, P.; Costanzo, P. KRAB-zinc finger proteins: A repressor family displaying multiple biological functions. *Curr. Genom.* 2013, 14, 268–278. [CrossRef] [PubMed]
- Alders, M.; Ryan, A.; Hodges, M.; Bliek, J.; Feinberg, A.P.; Privitera, O.; Westerveld, A.; Little, P.F.; Mannens, M. Disruption of a novel imprinted zinc-finger gene, ZNF215, in Beckwith–Wiedemann Syndrome. *Am. J. Hum. Genet.* 2000, *66*, 1473–1484. [CrossRef] [PubMed]