e) Basically, we need to state the conditions under which the transformation $\mathbf{T}$ complies with $p(\underline{x}_0 \mid N(\underline{\mu},\boldsymbol{\Sigma})) = p(\mathbf{T}\underline{x}_0 \mid N(\mathbf{T}^{\mathrm{T}}\underline{\mu}\ ,\ \mathbf{T}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{T}))$. Let's start by looking at the right hand side of the previous equation:

$$\frac{1}{(2\pi)^{d/2}\left|\boldsymbol{\Sigma}\right|^{1/2}}\exp\left[-\frac{1}{2}\left(\underline{x}-\underline{\mu}\right)^T\boldsymbol{\Sigma}^{-1}\left(\underline{x}-\underline{\mu}\right)\right]\xrightarrow[\substack{\underline{x}\to\mathbf{T}^T\underline{x}\\ \underline{\mu}\to\mathbf{T}^T\underline{\mu}}]{\boldsymbol{\Sigma}\to\mathbf{T}^T\boldsymbol{\Sigma}\mathbf{T}}$$

$$\frac{1}{(2\pi)^{d/2}\left|\mathbf{T}^T\boldsymbol{\Sigma}\mathbf{T}\right|^{1/2}}\exp\left[-\frac{1}{2}\left(\mathbf{T}^T\left(\underline{x}-\underline{\mu}\right)\right)^T\left(\mathbf{T}^T\boldsymbol{\Sigma}\mathbf{T}\right)^{-1}\left(\mathbf{T}^T\left(\underline{x}-\underline{\mu}\right)\right)\right]$$

Let's expand the exponent of the transformed Gaussian:

$$-\frac{1}{2}\left(\mathbf{T}^T\left(\underline{x}-\underline{\mu}\right)\right)^T\left(\mathbf{T}^T\boldsymbol{\Sigma}\mathbf{T}\right)^{-1}\left(\mathbf{T}^T\left(\underline{x}-\underline{\mu}\right)\right)\xrightarrow[\text{Use }(AB)^{-1}=A^{-1}B^{-1}]{\text{Use }(AB)^{T}=A^{T}B^{T}}-\frac{1}{2}\left(\underline{x}-\underline{\mu}\right)^T\mathbf{T}\left(\mathbf{T}^{-1}\boldsymbol{\Sigma}^{-1}\left(\mathbf{T}^T\right)^{-1}\right)\mathbf{T}^T\left(\underline{x}-\underline{\mu}\right)$$

Now, observe that the matrix multiplications associated with the transformation $\mathbf{T}$ all reduce to the identity matrix, which finally gives us the expression $-\frac{1}{2}\left(\underline{x}-\underline{\mu}\right)^T\boldsymbol{\Sigma}^{-1}\left(\underline{x}-\underline{\mu}\right)$, which is nothing more than what we had as the exponent of the original distribution. In order to complete this explanation, we also need to prove that $\left|\boldsymbol{\Sigma}\right|^{1/2}=\left|\mathbf{T}^T\boldsymbol{\Sigma}\mathbf{T}\right|^{1/2}$. A property of the determinants states that $\left|\mathbf{AB}\right|=\left|\mathbf{A}\right|\left|\mathbf{B}\right|$, so $\left|\boldsymbol{\Sigma}\right|=\left|\mathbf{T}^T\boldsymbol{\Sigma}\mathbf{T}\right|=\left|\boldsymbol{\Sigma}\right|\left|\mathbf{T}^T\mathbf{T}\right|$ is true only when $\left|\mathbf{T}^T\mathbf{T}\right|=1$. This means that $\mathbf{T}^T\mathbf{T}=\mathbf{I}$ or $\mathbf{T}^T=\mathbf{T}^{-1}$. This is a property that the orthogonal matrices have. We conclude that for $p(\underline{x}_0 \mid N(\underline{\mu},\boldsymbol{\Sigma})) = p(\mathbf{T}\underline{x}_0 \mid N(\mathbf{T}^{\mathrm{T}}\underline{\mu}\ ,\ \mathbf{T}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{T}))$ to be true, the transformation $\mathbf{T}$ must be orthogonal.

f) If the application of $\mathbf{A}_{\mathrm{w}}$ to a Gaussian distribution transforms its covariance matrix into the identity matrix, then the following must be true:

$$\frac{1}{(2\pi)^{d/2}\left|\mathbf{A}_{\mathrm{w}}{}^T\boldsymbol{\Sigma}\mathbf{A}_{\mathrm{w}}\right|^{1/2}}\exp\left[-\frac{1}{2}\left(\mathbf{A}_{\mathrm{w}}{}^T\left(\underline{x}-\underline{\mu}\right)\right)^T\left(\mathbf{A}_{\mathrm{w}}{}^T\boldsymbol{\Sigma}\mathbf{A}_{\mathrm{w}}\right)^{-1}\left(\mathbf{A}_{\mathrm{w}}{}^T\left(\underline{x}-\underline{\mu}\right)\right)\right]=$$

$$\frac{1}{(2\pi)^{d/2}\left|\mathbf{I}\right|^{1/2}}\exp\left[-\frac{1}{2}\left(\mathbf{A}_{\mathrm{w}}{}^T\left(\underline{x}-\underline{\mu}\right)\right)^T(\mathbf{I})^{-1}\left(\mathbf{A}_{\mathrm{w}}{}^T\left(\underline{x}-\underline{\mu}\right)\right)\right].$$

Which implies nothing but $\mathbf{A}_{\mathrm{w}}{}^T\boldsymbol{\Sigma}\mathbf{A}_{\mathrm{w}}=\mathbf{I}$. To prove this is true, we use the definition of the whitening transformation $\mathbf{A}_{\mathrm{w}}=\boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Phi}^T$ and get:

$$\mathbf{A}_{\mathrm{w}}{}^T\boldsymbol{\Sigma}\mathbf{A}_{\mathrm{w}}=\left(\boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Phi}^T\right)^T\boldsymbol{\Sigma}\left(\boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Phi}^T\right)$$

$$\left(\left(\boldsymbol{\Phi}^T\right)^T\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Phi}^T\right)\boldsymbol{\Sigma}\left(\boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Phi}^T\right)=\left(\boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Phi}^T\right)\boldsymbol{\Sigma}\left(\boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Phi}^T\right)$$

Matrix fundamentals tell us that $\mathbf{D}=\mathbf{X}^{-1}\mathbf{A}\mathbf{X}$ is a diagonal matrix, with the eigenvalues of $\mathbf{A}$ as entries if the columns of $\mathbf{X}$ are the eigenvectors of $\mathbf{A}$. Note that $\boldsymbol{\Phi}$ is a matrix which columns are the eigenvectors of $\boldsymbol{\Sigma}$ and because $\boldsymbol{\Sigma}$ is symmetric (covariance matrix) its eigenvectors form an orthonormal basis. That is, the matrix $\boldsymbol{\Phi}$ is orthogonal ($\boldsymbol{\Phi}^{-1}=\boldsymbol{\Phi}^T$). Now we can see that:

$$\mathbf{A_w}^T \Sigma \mathbf{A_w} = \Phi \Lambda^{-1/2} \underbrace{\Phi^T \Sigma \Phi}_{\Lambda} \Lambda^{-1/2} \Phi^T = \mathbf{I}$$

where Λ is a diagonal matrix with the eigenvalues of Σ as entries. Further development states:

$$\Phi \Lambda^{-1/2} \Lambda \Lambda^{-1/2} \Phi^T = \Phi \Lambda^{-1/2} \Lambda^{1/2} \Phi^T = \Phi \ \mathbf{I} \ \Phi^T = \mathbf{I}$$

The first part of the problem is checked. Now we have to check if the normalization is preserved by the $\mathbf{A}_{\mathrm{w}}$ transformation. We know from e) that only orthogonal transformations will keep normalization. In order to see if $\mathbf{A}_{\mathrm{w}}$ is orthogonal, we take the norm of the transformation. If it is equal to one, normalization is preserved (implying $\mathbf{A}_{\mathrm{w}}$ is orthogonal). Else, there will be a scaling factor equal to the norm of the transformation.

$$\left\| \mathbf{A_w}^T \mathbf{A_w} \right\| = \left\| \Phi \Lambda^{-1/2} \underbrace{\Phi^T \Phi}_{\mathbf{I}} \Lambda^{-1/2} \Phi^T \right\| = \left\| \Phi \Lambda^{-1} \Phi^T \right\| = \left\| \Lambda^{-1} \right\|$$

Finally we conclude that $\mathbf{A}_{\mathrm{w}}$ is not orthogonal, and does not preserve normalization. The whitening transform introduces a scaling factor of $\left\| \Lambda^{-1} \right\|$.