Part 2

① $p(\beta|\Sigma, X, y) = MVN\Big(\mathbb{E}[\beta|\Sigma, X, y] = Var[\beta|\Sigma, X, y]^{-1}(X^T\Sigma^{-1}y + \Sigma_\beta^{-1}\beta_0),$

$Var[\beta|\Sigma, X, y] = [X^T\Sigma^{-1}X]^{-1} + \Sigma_\beta^{-1}\Big)$

Substitute $\Sigma = \sigma^2 I$:

$$Var[\beta|\sigma^2 I, X, y) = [X^T(\sigma^2 I)^{-1}X]^{-1} + \Sigma_\beta^{-1}$$
$$= \left[\frac{1}{\sigma^2}(X^TX)\right]^{-1} + \Sigma_\beta^{-1} \quad \text{since } (\sigma^2 I)^{-1} = \frac{1}{\sigma^2}\cdot I$$
$$= \sigma^2(X^TX)^{-1} + \Sigma_\beta^{-1}$$

$$\mathbb{E}[\beta|\sigma^2 I, X, y] = Var[\beta|\sigma^2 I, X, y]^{-1}(X^T(\sigma^2 I)^{-1}y + \Sigma_\beta^{-1}\beta_0)$$
$$= Var[\beta|\sigma^2 I, X, y]^{-1}\left(\frac{1}{\sigma^2}X^Ty + \Sigma_\beta^{-1}\beta_0\right)$$
$$= (\sigma^2(X^TX)^{-1} + \Sigma_\beta^{-1})^{-1}\left(\frac{1}{\sigma^2}X^Ty + \Sigma_\beta^{-1}\beta_0\right)$$

$\therefore$, $p(\beta|\Sigma=\sigma^2 I, X, y) = MVN\Big(\mathbb{E}[\beta|\Sigma=\sigma^2 I, X, y] = (\sigma^2(X^TX)^{-1} + \Sigma_\beta^{-1})^{-1}\left(\frac{1}{\sigma^2}X^Ty + \Sigma_\beta^{-1}\beta_0\right),$

$Var(\beta|\Sigma=\sigma^2 I, X, y] = \sigma^2(X^TX)^{-1} + \Sigma_\beta^{-1}\Big)$

② If you're referring to the general case where Σ is free, then:

$$\mathbb{E}[\beta|\Sigma, X, y] = Var(\beta|\Sigma, X, y]^{-1}(X^T\Sigma^{-1}y + \Sigma_\beta^{-1}\beta_0)$$

If you're referring to $\Sigma = \sigma^2 I$, then:

$$\mathbb{E}[\beta|\Sigma=\sigma^2 I, X, y] = (\sigma^2(X^TX)^{-1} + \Sigma_\beta^{-1})^{-1}$$

③ To achieve $\mathbb{E}[\beta|\Sigma, X, y] = (X^TX)^{-1}X^Ty$, we need Σ_β^{-1} to be a zero matrix. β_0 can be free.

Since $\mathbb{E}[\beta|\Sigma, X, y] = Var[\beta|\Sigma, X, y]^{-1}(X^T\Sigma^{-1}y + \Sigma_\beta^{-1}\beta_0)$
$$= ([X^T\Sigma^{-1}X]^{-1} + \Sigma_{\beta_0}^{-1})^{-1}(X^T\Sigma^{-1}y + \Sigma_\beta^{-1}\beta_0)$$
$$= (X^T\Sigma^{-1}X)^{-1}(X^T\Sigma^{-1}y) \quad \Sigma_\beta^{-1} = 0$$
$$= (X^T(\sigma^2 I)^{-1}X)^{-1}(X^T(\sigma^2 I)^{-1}y) \quad \text{since } \Sigma = \sigma^2 I$$
$$= \sigma^2(X^TX)^{-1}\cdot\frac{1}{\sigma^2}(X^Ty)$$
$$= (X^TX)^{-1}(X^Ty)$$

④ $\mathbb{E}[\hat{y} = X\beta \mid \Sigma, X, y] = X\,\mathbb{E}[\beta \mid \Sigma, X, y]$, since X is fixed and

$$= X\left(\left([X^T\Sigma^{-1}X]^{-1} + \Sigma_{\beta_0}^{-1}\right)^{-1}\left(X^T\Sigma_y^{-1} + \Sigma_\beta^{-1}\beta_0\right)\right)$$

$$= X(X^TX)^{-1}X^Ty$$ we set Σ_β to be the zero matrix and get the same result as ③

$\therefore$, The hyperparameter specifications are the same as ③ (i.e. Σ_β is a zero matrix)

⑤ $Var(\beta \mid \Sigma = \sigma^2 I, X, y) = \left(\sigma^2 (X^TX)^{-1} + \Sigma_\beta^{-1}\right)^{-1}$